ISSN 2359-0386

4ª EDIÇÃO

ERI·GO

ESCOLA REGIONAL DE INFORMÁTICA

ANAIS

ANAIS

2016

Goiânia,GO

Organização do Material

Vinicius da Cunha M. Borges (UFG/Goiânia)
Antonio Carlos de Oliveira Júnior (UFG/Goiânia)
Luciana de Oliveira Berretta (UFG/Goiânia)

Editoração

Vinicius da Cunha M. Borges (UFG/Goiânia)

Dados Internacionais de Catalogação na Publicação (CIP)
GPT/BC/UFG

E75 Escola Regional de Informática de Goiás (04. : 2016 : Goiânia, GO).
Anais da IV Escola Regional de Informática de Goiás [recurso eletrônico] / Organizadores, Antonio Carlos de Oliveira Júnior, e Vinicius Cunha Martins Borges. – Dados eletrônicos. – Goiânia : Instituto de Informática, UFG, 2016.
320 p.

ISSN: 2359-0386
Disponível também em World Wide Web:
<http://erigo.sbc.org.br/up/4/o/anais-iv-erigo-2016.pdf>

1. Informática – Estudo e ensino. 2. Informática – Aspectos sociais e econômicos. I. Oliveira Júnior, Antonio Carlos de. II. Borges, Vinicius Cunha Martins. III. Título.

CDU: 004

COMITÊ

ORGANIZAÇÃO GERAL LOCAL

Antonio Carlos de Oliveira Júnior (UFG/Goiânia) - Coordenador
Marcelo Akira Inuzuka (UFG/Goiânia) - Vice-Coordenador

COORDENAÇÃO DO COMITÊ CIENTÍFICO

Vinicius da Cunha M. Borges (UFG/Goiânia) - Coordenador
Luciana de Oliveira Berretta (UFG/Goiânia) - Vice-Coordenadora

COMITÊ DE AVALIAÇÃO TÉCNICA E CIENTÍFICA

Adriano Braga (IFGoiano/Ceres)
Alessandro Silva (IFG/Anápolis)
Ana Claudia Loureiro Monção (UFG/Goiânia)
André Ribeiro (IFGoiano/Rio Verde)
Antônio Oliveira Jr (UFG/Goiânia)
Cristiane Bastos Rocha Ferreira UFG/Goiânia)
Christiane Borges (IFG/Luziânia)
Daniel Stefany Duarte Caetano (UFU)
Edison Morais (UFG/Goiânia)
Edmar Yokome (UEG/Santa Helena)
Edmundo Spoto (UFG/Goiânia)
Eliane Raimann (IFGoias/Jataí)
Eliomar Lima (UFG/Goiânia)
Fabian Cardoso (Fesurv/Rio Verde)
Fábio S. Marques (IFG/Goiânia)
Francisco Xavier (Fatesg)
Gilmar Arantes (UFG/Goiânia)
Hebert da Silva (UFG/Goiânia)
Iwens Sene Jr (UFG/Goiânia)
Luciana Berretta (UFG/Goiânia)
Luiz Henrique Rauber Rodrigues (Senac/Santa Cruz do Sul/RS)
Marcelo Akira Inuzuka (UFG/Goiânia)
Marcelo Ricardo Quinta (UFG/Goiânia)
Marcos Alves Vieira (IFGoiano/Iporá)
Marcos Aurélio Batista (UFG/Catalão)
Marcos Caetano (UnB)
Marcos Wagner de Souza Ribeiro (UFG/Jataí)
Nádia Silva (UFG/Goiânia)
Ole Peter Smith (IME/UFG)
Otávio Calaça Xavier (IFG/Goiânia)
Renato Bulcão Neto (UFG/Goiânia)
Ricardo Rocha (UFG/Catalão)
Rodrigo Cândido Borges (IFG/Inhumas)
Rogério Silva (IFG/Inhumas)
Roney Lima (IFG/Jataí)
Sergio Silva (UFG/Catalão)
Sérgio Teixeira de Carvalho (UFG/Goiânia)
Taciana Novo Kudo (UFG/Goiânia)
Tércio Alberto Santos Filho (UFG/Catalão)
Vinicius da Cunha M. Borges (UFG/Goiânia)
Vinicius Sebba Patto (UFG/Goiânia)
Waldir Moreira (Associação Fraunhofer Portugal Research, Porto)
Wendell Bento Geraldes (IFG/Luziânia)
William Ferreira (UFG/Goiânia)

REVISORES EXTERNOS

Anderson Soares (UFG/Goiânia)
Bruno Silvestre (UFG/Goiânia)
Deller James Ferreira (UFG/Goiânia)
Karina Rocha Gomes da Silva (UFG/EMC/Goiânia)
Kleber Vieira Cardoso (UFG/Goiânia)
Leandro Luís Galdino de Oliveira (UFG/Goiânia)
Lenice Miranda Alves (UFG/Goiânia)
Leonardo Andrade Ribeiro (UFG/Goiânia)
Marcia Rodrigues Cappelle Santana (UFG/Goiânia)
Rogério Salvini (UFG/Goiânia)
Ronaldo Martins da Costa (UFG/Goiânia)
Thierson Rosa (UFG/Goiânia)
Telma Woerle Lima Soares (UFG/Goiânia)
Wanderley Alencar (UFG/Goiânia)

E-mail de Contato: erigo@inf.ufg.br

DIRETOR DO INSTITUTO DE INFORMÁTICA - UFG

Eduardo Simões Albulquerque

Apresentação

A 4ª Edição da Escola Regional de Informática de Goiás (ERI-GO 2016) é um evento promovido pela UFG, IFGoiano e IFG. A ERI-GO 2016 será realizada em conjunto com o Fórum Goiano de Software Livre (FGSL) contribuindo assim para a maior integração entre a comunidade goiana de pesquisadores, usuários e profissionais atuantes na área de Software Livre (SL). Esta edição do evento tem como objetivos: estimular ações de pesquisa, ensino e extensão sobre informática em geral nas instituições de ensino superior e técnico como também centros de pesquisa no Estado de Goiás, baseado em trilhas dos principais cursos de Informática em Goiás, no Brasil e no mundo de acordo com ACM (Association for Computing Machinery), por exemplo, Sistema de Informação, Engenharia de Software, Ciência da Computação e Engenharia a fins; tornar público os projetos de SL desenvolvidos regionalmente; e promover a cooperação entre produtores e usuários de software livre visando ampliar o desenvolvimento tecnológico e a inovação em soluções de SL no Estado. A ERI-GO 2016 ocorreu no dia 19 de novembro de 2016, em Goiânia,GO. Vinte e Quatro artigos completos (apresentação oral) foram incluidos nos anais selecionados pelo comitê de avaliação técnica e científica para a ERI 2016, os quais tratam de vários temas de pesquisa e desenvolvimento tecnológico em Software Livre e Informática de forma geral. Além disso, o melhor artigo do evento foi selecionado e premiado, intitulado "Um Estudo de Caso para a Segmentação de Falanges em Radiografias Carpais", levando em conta a avaliação da apresentação e a avaliação dos revisores. Os anais também estão disponibilizados eletronicamente nos sites: http://www.inf.ufg.br/~erigo/4/o/anais-iv-erigo-2016.pdf e http://erigo.sbc.org.br/p/152-anais-eri-go.

Sumário

Mapeamento Sistemático da Literatura sobre o Estado da Prática do Processo de Engenharia de Requisitos de Software

Adailton F. Araújo[1], **Bruno N. Machado**[1], **Juliano L. Oliveira,** [1]
Almir F. Silva [2]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras Qd D Câmpus Samambaia
74001-970 – Goiânia – GO – Brasil

[2]Decisão Sistemas
Rua Uberaba Qd 77 Lt 9 Jd Luz 74915-440 – Goiânia – GO – Brasil

{adailton, brunonunes, juliano}@inf.ufg.br

{almir}@decisaosistemas.com.br

Abstract. *This paper presents a systematic mapping of studies related to topics of Software Requirements Engineering (SRE), specifically, about the practices in the SRE process, dated between 2008 and 2013. Guided by three research questions, this systematic mapping presents the following evidence generated from these studies: the main activities in the SRE process, the products generated by this process and the challenges for implementation of this process. The results suggest that there is still a huge lack of studies in some areas of SRE. As pointed out these niches unexplored, this mapping can serve as the basis and motivation for future researchers develop research in these areas still lacking.*

Resumo. *Este trabalho apresenta um mapeamento sistemático de estudos relacionados aos tópicos de Engenharia de Requisitos de Software (ERS), especificamente, a respeito das práticas no processo de ERS, datados entre 2008 e 2013. Orientado por três questões de pesquisa, este mapeamento sistemático da literatura apresenta as seguintes evidências geradas a partir desses estudos: as principais atividades no processo de ERS, os produtos gerados por esse processo e os desafios para execução desse processo. Os resultados sugerem que ainda há uma enorme carência de estudos em algumas áreas da ERS. Uma vez apontados estes nichos pouco explorados, este mapeamento poderá servir como base e motivação para futuros pesquisadores desenvolverem pesquisas nestas áreas ainda carentes.*

1. Introdução

A Engenharia de Requisitos (ER) está presente ao longo do ciclo de vida do software, identificando as expectativas das partes interessadas no sistema e definindo o que o sistema deve fazer para satisfazer essas expectativas e, embora seja uma disciplina relativamente nova, a ER já estabeleceu um conjunto de soluções comuns para descobrir, analisar, validar, documentar e gerenciar requisitos de software [Hull et al. 2011].

O Processo de Engenharia de Requisitos de Software (ERS) produz os requisitos que um determinado software deve atender. Esses requisitos são insumos para diversos

outros processos de software, tais como o Desenvolvimento, a Manutenção, os Testes e o Suporte Técnico à operação do software. Compreender a forma como os requisitos são identificados e registrados é essencial não apenas para a produção de software de qualidade, mas também para a utilização e a evolução deste software.

Nessa linha, este trabalho objetiva identificar o estado da prática do processo de ERS, que engloba identificar as principais atividades, produtos e desafios relacionados. Para alcançar tais objetivos, será aplicado o mapeamento sistemático como método de pesquisa para fornecer uma visão geral da área de ERS e permitir identificar a quantidade e o tipo de estudos e resultados disponíveis em tal área.

O restante deste trabalho está assim organizado: a Seção 2 traz todos as etapas para realização do mapeamento sistemático. A Seção 3 discute os resultados obtidos, respondendo a cada questão de pesquisa. Por fim, a Seção 4 traz conclusões finais do mapeamento

2. Mapeamento Sistemático

O mapeamento sistemático conduzido e apresentado neste trabalho está dividido em quatro fases: Planejamento, Condução, Seleção, Extração e Análise de Dados, de acordo com [Kitchenham and Charters 2007]. Uma breve descrição e os resultados dessas fases são apresentadas nas seções seguintes.

2.1. Planejamento

Nesta fase foi estabelecido o protocolo do mapeamento sistemático, onde foram definidas as questões de pesquisa, a *string* de busca, os critérios de inclusão, os critérios exclusão e os tipos de estudos que fizeram parte do escopo do mapeamento.

- Questões de Pesquisa

Com o objetivo de mapear o estado da prática da ERS, foram definidas as seguintes questões de pesquisas:

QP1. Quais são as principais atividades do processo de Engenharia de Requisitos de Software?

QP2. Quais são os principais produtos gerados pelo processo de Engenharia de Requisitos de Software?

QP3. Quais são os principais desafios (ou ameaças) para a execução do processo de Engenharia de Requisitos de Software?

- String de Busca

A *string* foi dividida em três blocos, separados pelo operador lógico "AND". O primeiro objetiva buscar trabalhos relacionados ao contexto geral da ERS, o segundo, trabalhos que estejam relacionados ao ciclo de vida da engenharia de requisitos e por fim, o objetivo do terceiro bloco é buscar aspectos de interesse dentro das atividades do ciclo de vida. A seguir são listadas as *strings* gerais utilizadas na busca dos estudos:

1. *String* em Inglês:

("Software Requirements" OR "Requirements Engineering")
AND

(Process OR Activity OR Modeling OR Elicitation OR Analysis OR Specification OR Validation OR Verification OR Management OR Reuse OR Documentation OR Evolution OR Life Cycle)
AND
(Product OR Framework OR Risk OR Challenge OR Threat OR Artifact OR Lesson OR Rastreability OR Agile OR Pattern OR Model OR Case OR Study OR Practice OR Experiment OR Empirical)

2. *String* em Português:

(“Requisitos de Software” OR “Engenharia de Requisitos”)
AND
(Processo OR Atividade OR Modelagem OR Elicitação OR Análise OR Especificação OR Validação OR Verificação OR Gerência OR Reuso OR Documentação OR Evolução OR Ciclo de Vida)
AND
(Produto OR Framework OR Risco OR Desafio OR Ameaça OR Artefato OR Lições OR Rastreabilidade OR Ágil OR Padrão OR Modelo OR Caso OR Estudo OR Prática OR Experimento OR Empírico)

- Seleção de Fontes de Pesquisa

Com o objetivo de obter um conjunto de fontes de buscas que atendesse as necessidades do mapeamento sistemático, foi elencado um conjunto de critérios de seleção de fontes. Com isso, para ser selecionada, uma fonte de busca deveria atender aos seguintes critérios:

1. Fornecer mecanismos de busca por meio de *strings* de busca.
2. Produzir os mesmos resultados sempre que a mesma *string* de busca for inserida.
3. Disponibilizar mecanismo de consulta via web.
4. Ser relacionada a temas de Computação, incluindo Ciência da Computação, Sistemas de Informação e/ou Engenharia de Software.
5. Permitir consulta online por meio de mecanismo de busca que permita expressões lógicas para definição de *string* de busca.
6. Permitir buscar simultaneamente apenas pelo título e resumo (abstract) dos estudos.
7. Permitir filtro por ano de publicação.
8. Não possuir limitação de tamanho que impossibilite a execução da string definida.

Para condução do mapeamento sistemático, foi necessário elencar quais fontes de busca seriam utilizadas. Para tal, iniciou-se com um conjunto de fontes de busca candidatas recomendadas pela comunidade acadêmica. Em seguida foram aplicados os critérios de seleção de fontes no conjunto inicial, a fim de obter um conjunto que atendesse os requisitos para condução do mapeamento. A Tabela 1 apresenta a lista de fontes de buscas que atenderam a todos os critérios e por isso foram selecionadas.

- Critérios de Inclusão de Estudos

Foram incluídos no mapeamento estudos publicados entre os anos 2008 e 2013, que tenham como base a realização de atividades do processo de ERS em um conjunto significativo de projetos de desenvolvimento ou manutenção de software.

Tabela 1. Fontes de busca utilizadas para o mapeamento

Fontes de Busca	Link
ACM	http://dl.acm.org/
IEEEXplorer	http://ieeexplore.ieee.org/
Scopus	http://www.scopus.com/
Web of Science	http://apps.webofknowledge.com
WER	http://wer.inf.puc-rio.br/
Engineering Village	http://www.engineeringvillage.com/
Science Direct	http://www.sciencedirect.com/
BDTD	http://bdtd.ibict.br/

- Critérios de Exclusão de Estudos

Foram excluídos do mapeamento estudos:

1. que não estavam escritos nos idiomas Português ou Inglês.
2. publicados antes de 2008.
3. publicados depois de 2013.
4. os estudos cujo texto completo não estivesse disponível para acesso gratuitamente na Web ou no portal de periódicos da Capes
5. publicados por meios que não exigem revisão por pares.
6. trabalhos não relacionados com a área de Engenharia de Requisitos.
7. sem evidências suficientes de validação empírica ou experimental das propostas.
8. que não discutissem atividades, produtos ou riscos do processo de ERS.

- Tipos de Estudos

Foram considerados os trabalhos do tipo tese, dissertação, capítulo de livro com *abstract* e artigo de periódico ou de conferência. Além disso, foram considerados estudos primários, secundários ou terciários, que satisfizessem os critérios de inclusão e que não se enquadrassem nos critérios de exclusão de estudos definidos neste trabalho.

2.2. Condução

Essa etapa consistiu na execução das *strings* de busca em cada uma das Máquinas de Buscas selecionadas e na preparação das informações retornadas por essas máquinas para execução das etapas seguintes. Para facilitar a avaliação e seleção dos trabalhos retornados pelas Máquinas de Buscas, as informações dos trabalhos foram armazenadas em uma planilha Excel. Assim, o produto gerado por essa etapa foi uma planilha Excel contendo as informações mais relevantes dos trabalhos retornados pelas máquinas de buscas, sem os duplicados e uma parametrização que permitiu ao pesquisador classificar e inserir mais informações relevantes para cada trabalho.

2.3. Seleção

Essa etapa consistiu na seleção dos artigos com base no resultados das buscas realizadas, através das *strings* submetidas nas máquinas de busca definidas no protocolo e apresentadas na Subseção 2.1.Esta primeira avaliação considerou o título e *abstract* de cada trabalho, tomando como parâmetros os critérios de inclusão e exclusão definidos na Subseção 2.1. Um artigo foi selecionado após ter passado pelos seguintes níveis de avaliação:

1. O Primeiro Pesquisador avaliou e indicou o status do estudo (I = Incluído, E = Excluído, D = Dúvida). Para os arquivos excluídos, atribuiu um dos critérios de exclusão. Para os arquivos incluídos , indicou quais questões de pesquisa estão relacionadas ao trabalho e informou, nas observações do trabalho, palavras chaves que indicassem esse relacionamento.
2. Um Segundo Pesquisador realizou a revisão dos estudos já avaliados pelo primeiro pesquisador, para confirmar ou indicar um status diferente para o estudo. Além disso, o segundo pesquisador também podia fazer ajustes nas observações incluídas pelo pesquisador 1.

2.4. Extração

Nesta etapa os pesquisadores realizaram a leitura do texto dos trabalhos na íntegra, e isso pôde esclarecer dúvidas e permitir uma decisão mais embasada de sua manutenção ou sua exclusão, por não satisfazer os critérios estabelecidos. Para isso, foi feito o download dos estudos incluídos na fase de Seleção, armazenados em um repositório pré-definido, os documentos foram nomeados com o título do trabalho e realizou-se a avaliação seguindo os critérios estabelecidos.

2.5. Análise de dados

Essa etapa consistiu na emissão de todos os relatórios e gráficos com base nos resultados encontrados por cada máquina de busca, status de cada artigo e demais atributos de cada artigo, que são discutidos e apresentados nas sessões seguintes.

Durante o processo de mapeamento dos estudos relacionados, a investigação foi conduzida com base nos critérios de inclusão e exclusão definidos na Subseção 2.1. Em algumas situações, os critérios de decisão de inclusão e/ou exclusão foram verificados e discutidos várias vezes entre a equipe de pesquisadores.

A varredura por estudos relevantes e confiáveis iniciou-se por meio da aplicação da *string* de busca (Subseção 2.1) nas fontes de busca apontadas na Subseção 2.1. Isso constituiu a etapa de condução do mapeamento (Subseção 2.2), a aplicação dos termos de busca em cada fonte de busca. O passo seguinte foi a coleta dos resultados apresentados em cada fonte de busca. Em seguida tais resultados (dados quantitativos) foram tabulados em uma planilha do MS Excel.

Neste mapeamento, os resultados obtidos foram refinados com base nos critérios de inclusão e exclusão (Subseção 2.1). Foram consideradas oito fontes de busca (Subseção 2.1). Como primeiro resultado, foi obtido um grande número de estudos, porém ao verificar a duplicidade dos estudos apontados em fontes diferentes, por meio de funcionalidades do próprio MS Excel, reduziu-se quase pela metade a quantidade de estudos obtidos. Mesmo assim o volume de trabalhos ainda era muito alto.

Em seguida deu-se início a etapa de seleção (Subseção 2.3), nesta etapa os resultados foram refinados avaliando o título e o resumo dos estudos. Assim, obteve-se mais uma redução no número de estudos. Vale lembrar, que a etapa de seleção foi realizada em dois níveis de análise por dois pesquisadores diferentes. Numa ultima etapa, na extração (Subseção 2.4), esses resultados foram refinados novamente por meio da análise na íntegra de cada estudo, apoiado pelos critérios de exclusão e inclusão. Assim, obteve-se o resultado final do mapeamento sistemático.

A Figura 1 apresenta o fluxo simplificado do processo de mapeamento sistemático realizado nesta pesquisa, com os valores referentes ao número de estudos resultantes em cada etapa, depois da aplicação dos critérios de inclusão e exclusão. As máquinas de buscas retornaram 6570 trabalhos e ao final, após a execução das etapas de seleção e extração, restaram 523 trabalhos.

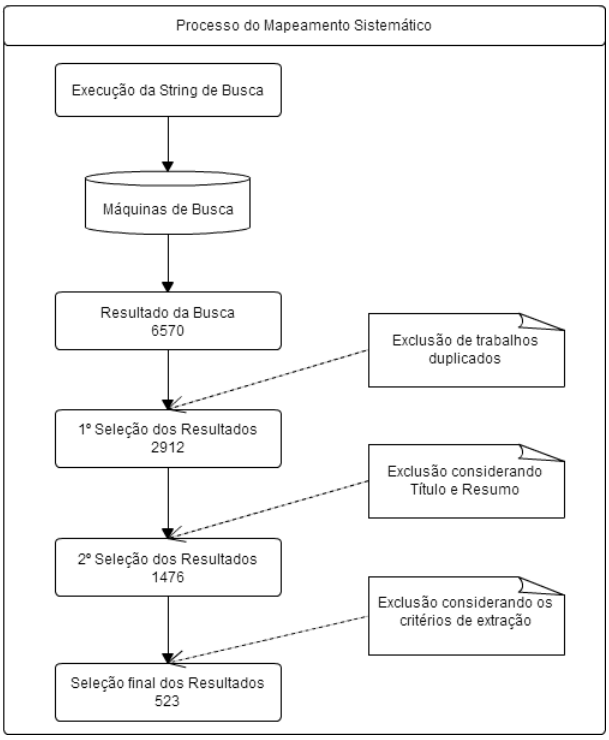


Figura 1. Processo de Revisão dos Estudos

Com foco na aplicação dos critérios de exclusão a Figura 2 mostra a quantidade de estudos excluídos por cada critério de exclusão na etapa de extração. O panorama apresentando na Figura 2 engloba um subconjunto de 953 estudos, dentro do total de 1476 estudos resultantes da etapa de seleção. Nesse subconjunto, é apresentado o percentual dos critérios que foram aplicados para excluir os 953 estudos. Nota-se que maioria dos estudos foram excluídos por não apresentarem validação da sua proposta, precisamente 453 estudos, o correspondente a 47.5% dos estudos excluídos.

Os estudos foram considerados sem validação quando estes não apresentavam evidências suficientes de validação empírica ou experimental das suas propostas. Além disso, considerou-se que a validação deveria ter ocorrido na industria e/ou na academia, porém neste último caso, a nível de pós-graduação. Isso resultou num grande número de estudos excluídos, pois uma grande fatia do conjunto de estudos analisados, não apresentavam a realização da validação, seja por meio de experimentação ou avaliação empírica.

Outro grande fator de exclusão foi a disponibilidade dos estudos, seja ela gratuita ou não. Na fase de extração, 271 estudos foram excluídos ou por não estarem realmente disponíveis em meio digital ou por serem pagos mesmo utilizando conta estudantil oferecida pela Universidade Federal de Goiás. Em seguida, a capacidade de responder as questões de pesquisa, foi outro critério que levou a exclusão de um número expressivo

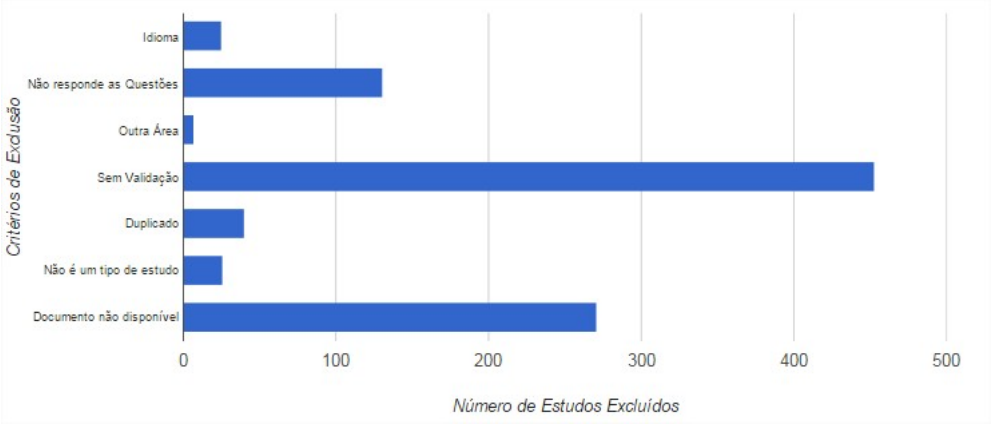


Figura 2. Motivos de Exclusão dos Estudos

de estudos. Na extração 131 estudos foram excluídos por não responderem as questões de pesquisa, ou seja, tais estudos não discutiam nenhuma prática da ERS, e também não apresentavam nenhum método ou ferramenta para apoiar o processo de ERS.

Além disso, foram excluídos 40 estudos por serem considerados duplicados. Durante a extração verificou-se que um mesmo autor poderia ter publicado a sua proposta em várias etapas do desenvolvimento de sua pesquisa, foram identificados estudos com propostas iniciais, ainda sem resultados, estudos completos com a proposta, resultados e discussões e ainda dissertações que englobavam ambos os estudos publicados. Nesses casos, considerou-se os artigos com propostas iniciais como sendo duplicados em relação aos que apresentam mesma proposta, porém com resultados e discussões. Em alguns casos, ambos os artigos foram considerados duplicados, e optou-se por permanecer com a dissertação, pois esta discutia de maneira mais aprofundada a proposta e os resultados obtidos.

A Subseção 2.1 indica quais são os tipos de estudos aceitos neste mapeamento. Nesse sentido, o critério "não é um tipo de estudo" enfatiza que serão aceitos apenas o subconjunto de estudos definidos na Subseção 2.1. Assim, foram excluídos 26 estudos aplicando este critério. Em seguida, o critério "Idioma", excluiu 25 estudos. Tais estudos passaram pelo processo de seleção apresentando título e *abstract* em inglês, porém na etapa de extração, os estudos estavam escritos em outros idiomas, muitas vezes em alemão, espanhol e também foram encontrados muitos estudos em idiomas orientais.

Por fim, o critério "Outra Área" excluiu 8 estudos. Foram considerados de outras áreas, estudos que não tratavam especificamente da engenharia de requisitos. Logo estudos que travam de engenharia de software em geral foram considerados outra área e excluídos. Aqui foi apresentado uma visão geral da condução do mapeamento e a aplicação dos critérios de exclusão, a seção seguinte apresentará um análise a cerca dos 523 estudos resultantes do mapeamento em relação a cada questão de pesquisa.

3. Discussão dos Resultados

Esta seção analisa os resultados obtidos do mapeamento sistemático sobre o processo de ERS, visando responder as questões de pesquisa apresentadas neste trabalho. Assim, a análise está subdividida, de forma a responder cada questão pesquisa (Subseção

2.1).

A análise considera um repositório com 523 estudos resultantes do processo de mapeamento sistemático apresentado neste trabalho e foca em duas perspectivas relacionadas a essa questão: a) estudos que apresentaram alguma resposta para as questões de pesquisa; e b) tais estudos agrupados por categoria.

3.1. Análise da Questão 1

A questão de pesquisa 1, visa identificar quais são as principais atividades do processo de ERS. Os estudos avaliados foram agrupados de acordo com as categorias de requisitos propostas em [ISO/IEC/IEEE 2009]. Neste mapeamento, cada categoria segue os seguintes conceitos:

A Eliciação de Requisitos identifica os *stakeholders* envolvidos no Domínio do Problema e descreve suas expectativas e necessidades com relação ao software que é objeto do processo de ERS [Burnay et al. 2014]. A Análise de Requisitos define modelos para a solução de software que atende as necessidades identificadas [ISO/IEC/IEEE 2011]. A Especificação de Requisitos documenta e apresenta, de diferentes maneiras, informações sobre os requisitos relevantes para diferentes stakeholders [IEEE 1998].

Verificação e Validação estão relacionadas à qualidade dos requisitos. A Verificação avalia a viabilidade técnica de um conjunto de requisitos enquanto a validação registra o comprometimento dos stakeholders com esses requisitos [ISO/IEC/IEEE 2011]. A Gerência de Requisitos estabelece baselines de requisitos e controla mudanças nessas baselines com base em informações de rastreabilidade de requisitos [ISO/IEC/IEEE 2011]. Além disso, foi elencada também a categoria Outros, para os casos em que os estudos apontassem atividades que não se encaixavam em nenhuma das categorias descritas aqui. Com foco apenas nestas categorias, a Figura 3 traz o percentual de trabalhos que apresentaram alguma proposta dentro dessas categorias.

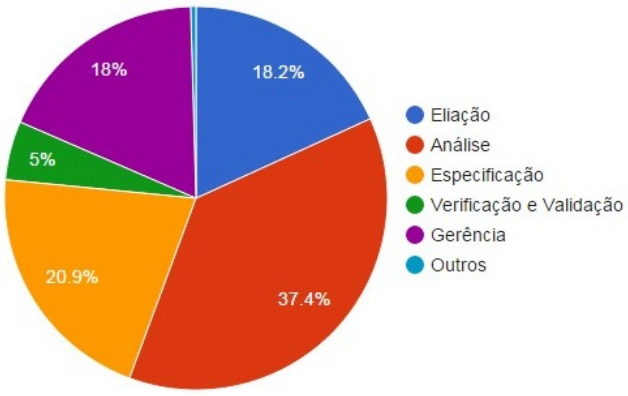


Figura 3. Estudos agrupados por Categoria

A Figura 3 mostra, para cada categoria da ER, o percentual de estudos avaliados que fizeram alguma proposta dentro de cada categoria. Por exemplo, dentre os 523 estudos resultantes, 37.2% apresentaram propostas relacionadas a categoria de Análise. Porém, isso não significa que um estudo que fez uma proposta em análise não possa ter

feito também uma proposta em eliciação, por exemplo. Logo, pode-se ter estudos que foram contabilizados em mais de uma categoria.

Esses resultados evidenciam que entre o período de 2008 à 2013, houve maior número de estudos publicados com foco em atividades relacionadas à análise e especificação do que para o restante das outras categorias. A verificação e Validação, por exemplo, foi contemplada por apenas 4.9% das propostas apresentadas entre 2009 e 2013.

Ao analisar cada categoria percebe-se, que determinadas atividades tem um número de proposta muito maior que outras, ou seja, pode-se dizer que os esforços da comunidade de ERS, estão focados em determinadas atividades de cada categoria. A Figura 4 traz a distribuição das atividades, agrupadas por cada categoria, que foram avaliadas durante o processo de mapeamento. As atividades apontadas para avaliação neste trabalho são propostas por [Bourque and Fairley 2014].

Nessa perspectiva, é possível notar que dentro da categoria Análise que teve 37.2% das propostas encontradas nos estudos conforme mostra a Figura 3, tais propostas estão focadas na atividade de modelagem. Nesta categoria, por exemplo, dos 250 estudos que apontaram proposta em alguma das atividades da análise, 157 fizeram alguma proposta relacionada a atividade de modelagem, ou seja, mais da metade dos trabalhos apontados nesta categoria realizaram propostas para a atividade de modelagem. O mesmo acontece na Eliciação, onde é possível notar um foco maior nas técnicas de eliciação, com 109 estudos, em relação ao tratamento das fontes de requisitos, com apenas 12 estudos.

A categoria Outros foi contemplada apenas com 0,45% dos estudos, mais precisamente, apenas 3 estudos apontaram atividades que não estavam alinhadas com atividades elencadas nas demais categorias. Tais estudos apresentavam propostas relacionadas as atividades de design, gerenciar riscos e ranquear requisitos.

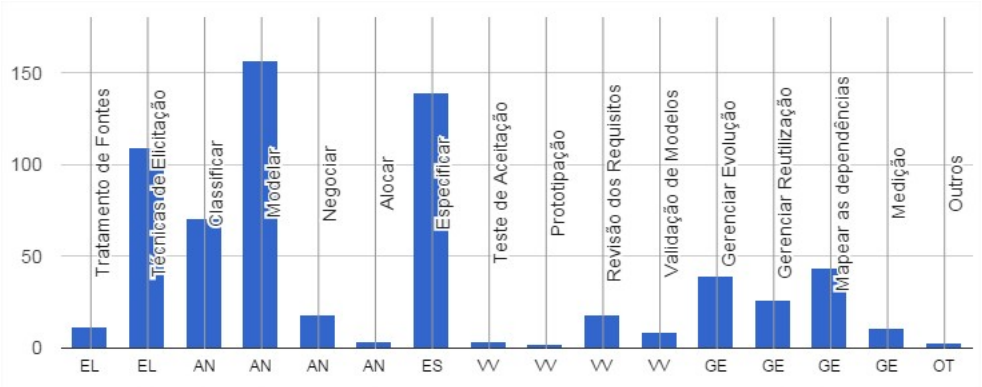


Figura 4. Estudos classificados por Atividades agrupadas por Categoria

Este cenário evidencia um grande foco da comunidade em atividades especificas de cada categoria, como por exemplo, a atividade de modelar na categoria de análise, que obteve maior concentração de propostas. No entanto, tal cenário não significa que as outras atividades não apresentem problemas a serem solucionados. A falta de estudos que contemplem de forma efetiva todas as atividades e categorias do processo de ERS, pode indicar que ainda há várias questões em aberto que requerem soluções ainda não apontadas.

3.2. Análise da Questão 2

A QP2 objetiva encontrar quais são os principais produtos gerados pelo processo de ERS. O conjunto de produtos elencados neste trabalho foi construído a partir da própria análise dos estudos identificados no mapeamento sistemático. Iniciou-se com um subconjunto pré-definido de produtos típicos gerados pelo processo de ERS, e durante a análise dos estudos resultantes tais produtos foram sendo confirmados e estendidos na etapa de extração. O conjunto final estabelecido abrange os resultados esperados apontados pela norma ISO/IEC/IEEE 29148 [ISO/IEC/IEEE 2011], apesar dela não apontar os produtos gerados após cada atividade, ela apresenta os resultados esperados de cada etapa do processo de ERS. Nessa linha, entende-se que tais resultados precisam ser documentos em produtos, e os produtos estabelecidos neste trabalho mapeiam ou cobrem os resultados esperados apontados em [ISO/IEC/IEEE 2011].

A Figura 5 mostra, dentre um conjunto de produtos possíveis de serem gerados pelo processo de ERS, o percentual de estudos avaliados que fizeram alguma proposta relacionada há alguns desses produtos. O panorama apresentando na Figura 5 engloba um subconjunto de 256 estudos, dentro do total de 523 estudos resultantes do mapeamento. Dentre o subconjunto de produtos avaliados, é possível notar uma tendência em encontrar propostas voltadas para algum tipo de diagrama. A Figura 5 mostra que dentre os estudos resultantes, 31.3% apresentaram propostas relacionadas ao produto Diagrama. Porém, isso não significa que um estudo que fez uma proposta em digrama não possa ter feito também uma proposta em protótipo, por exemplo. Logo, assim como nos estudos que tratavam de atividades, pode-se ter o mesmo estudo sendo contabilizado em mais de um produto.

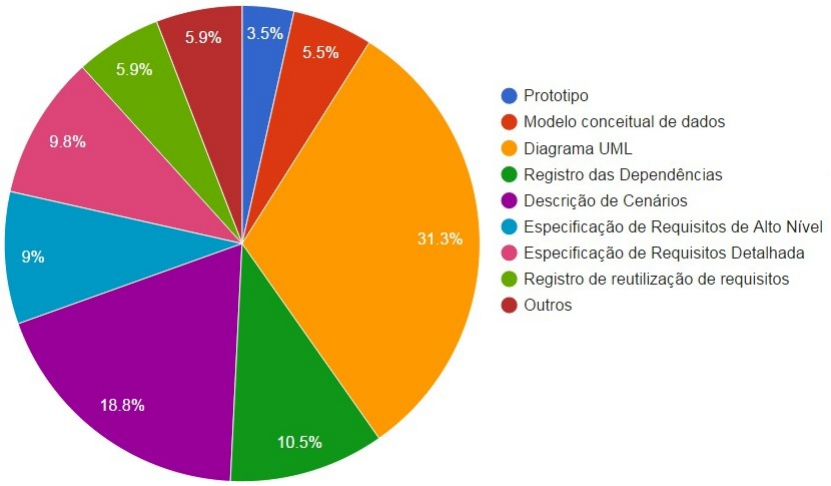


Figura 5. Produtos gerados pelo processo de ERS

A categoria Outros representa os produtos que foram apontados no estudos analisados e que não se encaixavam nas categorias já definidas. Dentre os produtos classificados como Outros, destaca-se o produto "questionários" que poderiam ser aplicados para apoiar a técnica de eliciação por meio questionários. Destaca-se também o produto "checklist", que por sua vez poderia ser um checklist para apoiar as atividades da categoria de verificação e validação. Dentre esses e outros produtos, a categoria Outros obteve 5.9%

do subconjunto de 256 estudos que apontaram algum produto que poderia ser gerado pelo processo de ERS.

Esses resultados evidenciam que entre 2008 e 2013, houve maior número de estudos publicados com foco em produtos relacionados há algum tipo de diagrama. Os protótipos foram os produtos que tiveram menos propostas dentre os estudos analisados, apenas 3.5% dos estudos apresentaram alguma proposta em relacionada à protótipos.

Ao analisar os resultados agrupados por categoria, percebe-se que a maioria dos estudos que abordam os possíveis produtos gerados pelo processo de ERS estão concentrados na categoria de análise. Dentro dessa categoria, o produto mais discutido nos estudos continuou sendo os diagramas, com 80 estudos que apontam algum tipo de diagrama.

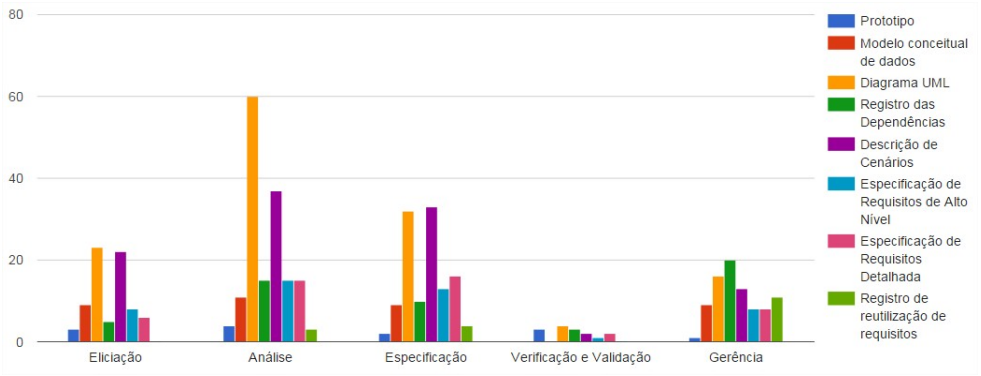


Figura 6. Produtos gerados pelos processo de ERS agrupados por Categoria

3.3. Análise da Questão 3

A QP3 objetiva encontrar quais são os principais desafios para a execução do processo de ERS. O conjunto de desafios elencados neste trabalho foi construído a partir da própria análise dos estudos identificados no mapeamento sistemático. Iniciou-se com um subconjunto pré-definido de desafios transversais às categorias do processo de ERS, e durante a análise dos estudos resultantes tais desafios foram sendo confirmados e estendidos na etapa de extração. O conjunto final estabelecido abrange apenas os aspectos transversais às categorias da ER, não considera-se os desafios naturais e com aspectos ortogonais de cada categoria de ER.

A Figura 7 mostra o percentual de estudos avaliados que apontaram desafios para execução do processo de ERS. O panorama apresentando na Figura 7 engloba um subconjunto de 557 estudos, dentro do total de 523 estudos resultantes do mapeamento. Vale lembrar que um estudo que fez uma proposta em análise não possa ter feito também uma proposta em eliciação, por exemplo. Logo, pode-se ter estudos que foram contabilizados em mais de uma categoria.

Dentre o subconjunto de desafios avaliados, é possível notar que a grande maioria dos estudos apontaram desafios relacionados a representação dos requisitos e a comunicação no processo de ERS, com 29.3% e 23% respectivamente. Vale ressaltar, que um mesmo estudo pode ter apontado mais de um desafio, logo este estudo foi contato para cada desafio apresentado.

Além dos desafios elencados, identificou-se estudos que tratavam de desafios distintos dos já elencados, tais desafios foram classificados como Outros. Nesta classificação, 3.8% dos estudos apontaram outros desafios transversais às categorias da ERS que não obtiveram um número representativo de estudos que tratassem de um determinado desafio. Dentre os desafios desse conjunto, encontra-se desafios como a ambiguidade de requisitos, a gestão do conhecimento dentre outros que não obtiveram um número de estudos representativos.

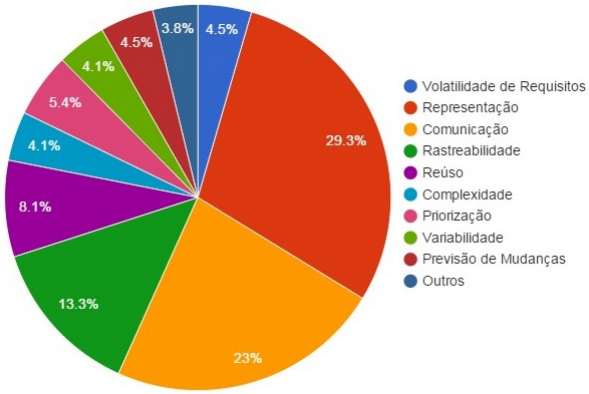


Figura 7. Desafios para execução do processo de ERS

Ao analisar os estudos que apontaram algum desafio agrupados por categoria (Figura 8), nota-se que a categoria de análise é a que apresentou o maior número de estudos com desafios apontados. Nessa categoria também é onde se tem a maior concentração de estudos relacionados aos desafios de representação de requisitos e comunicação no processo de ERS, que foram os mais apontados no contexto geral.

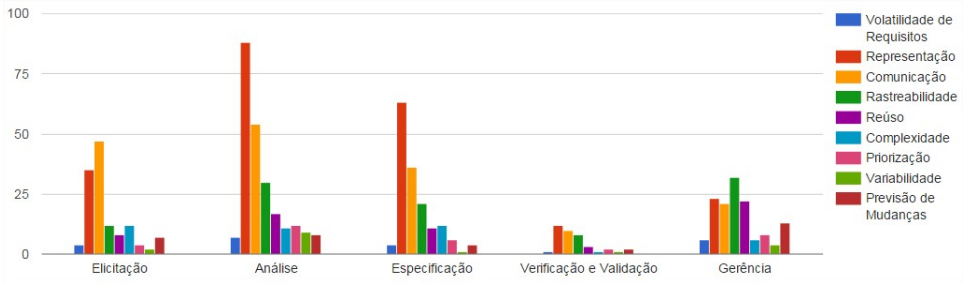


Figura 8. Desafios para execução do processo de ERS agrupados por Categoria

Este cenário evidencia que os desafios representação e comunicação no processo de ERS foram os desafios mais atacados pela comunidade entre o período de 2008 e 2013. Além disso, é possível notar também, que a categoria de análise foi a que mais apresentou desafios e a que mais foi atacada pela comunidade de ERS. No entanto, tal cenário não significa que não haja mais desafios na categiria de análise ou que não haja mais problemas na representação de requisitos e na comunicação.

4. Conclusões

Este trabalho apresentou um mapeamento sistemático da literatura que investigou quais são as práticas do processo de ERS, contemplando suas atividades, produtos gerados e desafios apontados. Por meio dos resultados obtidos neste mapeamento sistemático,

pode-se dizer que o estado da prática, considerando as atividades, os produtos e os desafios, dentro do processo de ERS ainda se encontra com grandes lacunas e possui bastante espaço para investigação.

O resultado deste mapeamento indica, ainda, que existe uma enorme carência de estudos sobre algumas etapas da ERS, já que a maioria dos estudos focaram nas categorias de Análise e Especificação. Encontrou-se poucos estudos relacionados à metodologias e/ou técnicas de verificação e validação, tanto para os resultados em inglês quanto para os em português. Mesmo com sua significativa importância para satisfazer as expectativas das partes interessadas, ainda assim as pesquisas nestas áreas estão carentes. Tal fato pode indicar ainda uma deficiência inerente às habilidades dos pesquisadores de engenharia de requisitos.

Além disso, a número de estudos que apresentaram evidências de validação empírica ou experimental de suas propostas, foi muito baixo, considerando o total de estudos levantados. Este cenário evidência a necessidade de adoção de práticas qualitativas, ou seja, é necessário destacar a importância de se realizar experimentação na engenharia de requisitos.

Uma vez apontados estes nichos pouco explorados, este mapeamento poderá servir como base e motivação para futuros pesquisadores desenvolverem pesquisas nestas áreas ainda carentes. Como trabalhos futuros, destaca-se: a) a realização da análise qualitativa sobre a base de conhecimento de 523 trabalhos gerada por este trabalho;e b) a seleção e avaliação de ferramentas de software que possam apoiar o processo de ERS.

Referências

[Bourque and Fairley 2014] Bourque, P. and Fairley, R. E., editors (2014). *SWEBOK: Guide to the Software Engineering Body of Knowledge*. IEEE Computer Society Press, Los Alamitos, CA, version 3.0 edition.

[Burnay et al. 2014] Burnay, C., Jureta, I. J., and Faulkner, S. (2014). What stakeholders will or will not say: A theoretical and empirical study of topic importance in requirements engineering elicitation interviews. *Information Systems*, 46(1):61–81.

[Hull et al. 2011] Hull, E., Jackson, K., and Dick, J. (2011). *Requirements engineering*. Springer, 3rd edition.

[IEEE 1998] IEEE (1998). *IEEE Standard 830-1998 - Recommended Practice for Software Requirements Specifications*. IEEE.

[ISO/IEC/IEEE 2009] ISO/IEC/IEEE (2009). *TR 24766: Information technology - Systems and software engineering - Guide for requirements engineering tool capabilities*. ISO.

[ISO/IEC/IEEE 2011] ISO/IEC/IEEE (2011). *Standard 29148-2011 - Systems and software engineering - Life cycle processes - Requirements Engineering*. ISO.

[Kitchenham and Charters 2007] Kitchenham, B. and Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering.

Uma revisão sobre métodos de desenvolvimento de aplicações móveis progressivas através de tecnologias para web

Gabriel de Lima Rabelo[1], Marcelo Ricardo Quinta[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal – 131 - CEP 74.690-900 – Goiânia – GO – Brasil

{gabrielrabelo,marcelo}@inf.ufg.br

Abstract: *The traffic of information on mobile browsers has overcome in 2015 the traffic on installed apps on mobile devices and, in the year of 2016, has also overcome the traffic on desktop browsers. However, the experience on mobile websites is still inferior to the experience on native apps installed on the device. By comprehending this problem and the potential of the new solutions available, this paper presents a review on the development methods of progressive mobile apps through web technologies which result on mobile websites that deliver a more complete and complex experience.*

Resumo: O tráfego de informações em navegadores móveis superou em 2015 o tráfego dentro dos aplicativos instalados em dispositivos móveis e, no ano de 2016, superou também o tráfego em navegadores desktop. Mesmo assim, a experiência em sites móveis ainda é inferior a aplicativos nativos instalados nos aparelhos. Conhecendo este problema e os potenciais das novas soluções disponíveis, este trabalho apresenta uma revisão sobre os métodos de desenvolvimento de aplicações móveis progressivas através de tecnologias para web que tem como resultado sites móveis que entreguem experiências mais complexas e completas.

1. Introdução

Desde outubro do ano de 2016, há mais acesso a internet por navegadores de dispositivos móveis do que em navegadores de computadores convencionais [Gibbs 2016]. Ao mesmo tempo, sites tem problemas ao prover uma experiência móvel mais completa quando precisam trabalhar com contextos comuns, porém problemáticos. Exemplos destes contextos são a falta de conectividade de internet e a necessidade de carregamento de bancos de dados complexos dentro do aparelho dos usuários.

Por outro lado, mesmo com menos tráfico comparativamente ao navegador móvel [Koetsier 2015] e com custo maior de desenvolvimento, os aplicativos instalados e desenvolvidos com tecnologias nativas dos sistemas operacionais móveis possuem acesso a mais ferramentas e recursos do sistema, facilitando a criação de uma aplicação mais completa e que seja mais confiável em relação a garantia de funcionamento de todas suas funções. Existiria então uma opção que tornasse o produto do desenvolvimento web mais próximo do que é entregue com aplicativos nativos, de forma que a experiência na internet fosse a mesma de um app?

Uma das respostas que o mercado atual alardeia é o conceito de "*progressive web apps*", uma metodologia de desenvolvimento em que aplicações são construídas com tecnologias para a *web*, ou seja, começam como sites móveis. Daí em diante, as aplicações são evoluídas até o momento que são empacotados como aplicativos a serem instalados nos aparelhos.

Sabendo deste novo conceito, este trabalho foi realizado com o propósito de

analisar o estado da arte do desenvolvimento de sistemas web e aplicativos através de tecnologias web, assim como caracterizar a técnica dos *progressive web apps*, fazendo um paralelo entre problemas e possíveis soluções vindas desta metodologia.

2. Metodologia e execução

Como este trabalho foi desenvolvido com o intuito de analisar o estado da arte de métodos atuais e buscar a fuga de interpretações parciais do mercado, foi necessário o uso de uma técnica que pudesse cobrir uma verificação mais completa sobre o atual contexto e desse maior assertibilidade, tanto na visão acadêmica quanto na do mercado. Por isso, a metodologia de trabalho adotada neste trabalho é a revisão sistemática segundo o protocolo de [Kitchenham 2004], dividida em cinco etapas, descritas nas seções seguintes:

1. **Planejamento**: Definição do protocolo de revisão,
2. **Execução:** Realização das buscas de material através de termos pré-definidos, em fontes anteriormente planejadas.
3. **Seleção:** Filtragem através de leitura do título e resumo dos materiais coletados, seguindo o protocolo de exclusão/duplicação previamente definido.
4. **Extração:** Leitura dos artigos previamente escolhidos.
5. **Análise e conclusão:** Divulgação de resultados das respostas (ou ausência destas) das questões de pesquisa, de acordo

2.1. Planejamento

No estágio de planejamento, o protocolo de revisão foi elaborado de maneira que definisse os objetivos finais e o meio pelo qual chegaríamos aos resultados. As questões de pesquisa são descritas a seguir:

- QP1: Quais são os problemas da web e aplicativos nativos hoje?
- QP2: Quais são as características esperadas de um *progressive web app*?
- QP3: Quais problemas atuais da web e de aplicações nativas podem ser resolvidos com *progressive web apps*?

As bases de conhecimento escolhidas foram, ACM, IEEE Xplore, Science Direct e SCOPUS. Dentro destas, foi utilizada uma String de busca contendo as palavras-chave nos campos de pesquisa avançada, definidas após avaliação com profissionais da área e referências existentes em artigos previamente conhecidos. As palavras-chave levantadas como referência ao tema foram: *progressive web applications*, *hybrid mobile applications*, *responsive web design*, *client-side web applications*, *progressive enhancement* e *native mobile applications.* Como resultado, a seguinte expressão foi aplicada em cada base: *"((((("progressive web applications") OR "hybrid mobile applications") OR "responsive web design") OR "client-side web applications") OR "progressive enhancement") OR "native mobile applications")"*.

Segundo o planejamento, um plano alternativo sobre a falta de material acadêmico sobre assunto foi a busca também pelo buscador *Google* com a mesma String, avaliando os resultados dentre as duas primeiras páginas da busca.

2.2. Execução

A execução da String de busca sobre as bases citadas retornou 558 resultados, como pode ser visto na Figura 1, que relaciona a quantidade de artigos e a distribuição percentual de resultados entre as bases de pesquisa. A Figura 2 mostra os resultados agrupados por ano de publicação. A execução da String no *Google* retornou 20 resultados.

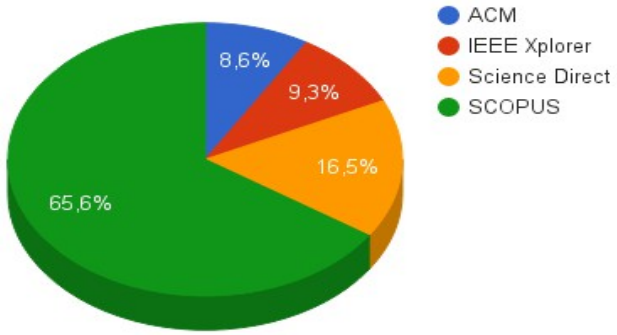


Figura 1 - Distribuição percentual de resultados por base de pesquisa.

2.3. Seleção

Este trabalho objetivou selecionar estudos que apresentaram informações relevantes para o estudo e desenvolvimento de aplicações web, móveis nativas e híbridas, bem como informações sobre novas soluções disponíveis para a construção de *progressive web apps*, com o intuito de levantar dados suficientes para a resolução das questões de pesquisa. Na fase de seleção foram avaliados o título e o resumo dos resultados das buscas nas bases de pesquisa e, com base nestes, foram definidos critérios de inclusão e exclusão dos resultados para filtragem. Foram também contabilizados os resultados duplicados entre as bases.

Critérios de inclusão:

1. O resultado trata de assuntos relacionados a *progressive web apps*.
2. O resultado apresenta definições e/ou contribuições sobre design responsivo.
3. O resultado apresenta definições e/ou contribuições sobre *progressive enhancement*.
4. O resultado apresenta desafios no desenvolvimento de aplicações web, móveis nativas e híbridas.
5. O resultado apresenta soluções para problemas atuais de desenvolvimento de aplicações web, móveis nativas e híbridas.

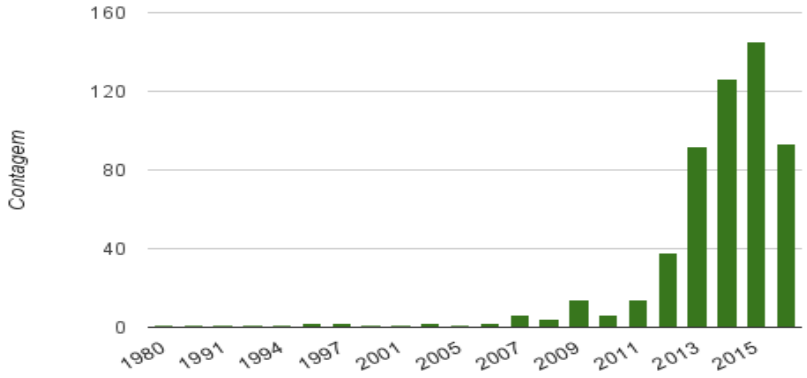


Figura 2 – Resultados agrupados por ano de publicação

Critérios de exclusão:

1. O resultado não está no formato PDF;
2. O resultado não está nas línguas inglesa ou portuguesa;
3. O resultado não está disponível online integralmente;
4. O resultado apenas menciona os termos pesquisados sem apresentar definições e outros aspectos técnicos sobre o assunto.
5. O resultado não é um artigo científico.

Para os resultados provenientes da busca no *Google* foram utilizados todos os critérios de inclusão descritos acima. Um único critério de exclusão foi definido:

1. O resultado não apresenta contribuições suficientes sobre o tema.

Como resultado da aplicação dos critérios de exclusão sobre os 558 resultados provenientes das bases de pesquisa, foram excluídos 418 artigos, juntamente com 93 duplicações. 5.37% dos resultados foram excluídos por não serem artigos científicos e apenas 2 resultados foram excluídos não estarem nas línguas inglesa e portuguesa. Não foram utilizados outros critérios de qualidade. A Figura 3 mostra uma distribuição dos resultados por base de pesquisa desta fase de seleção. Dos 20 resultados do *Google*, 19 foram incluídos e houve apenas 1 duplicação.

2.4. Extração

Na fase de extração foram lidos na íntegra todos os artigos e páginas incluídos na fase de seleção. A avaliação realizada nesta etapa considerou a qualidade do texto de cada item e sua contribuição em relação ao tema. Os artigos incluídos foram escolhidos com base nas questões de pesquisa. Como resultado, 19 artigos e 1 página foram extraídas.

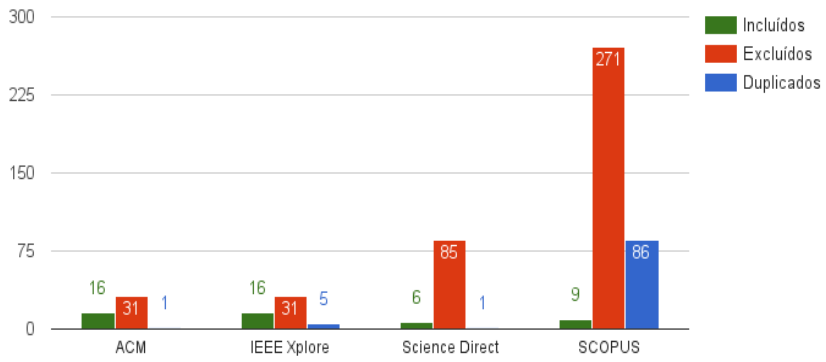


Figura 3 - Resultado do processo de seleção de artigos pelas bases de pesquisa.

As exclusões deram-se em casos em que o texto não apresentou informações aprofundadas ou relevantes em relação às palavras-chave utilizadas na fase de execução. Houveram também textos em que, após comparações com outros trabalhos, foram considerados que suas contribuições eram semelhantes a outros textos previamente extraídos.

2.5. Análise

Dada a leitura dos artigos e páginas resultantes, apresentamos nesta seção as respostas às questões de pesquisa levantadas anteriormente, sendo importante destacar que nem todos trabalhos respondem a todas. A compreensão do conjunto foi o que nos permitiu responder cada pergunta.

QP1: Qual problema da web e aplicativos nativos hoje?

Foram constatados diversos pontos desafiadores para o processo de desenvolvimento de aplicações web e móveis nativas. Estão descritos a seguir alguns dos pontos mencionados pelos autores dos artigos escolhidos.

Em relação à web:

1. **Heterogeneidade de navegadores**: Não existe um acordo entre os navegadores em relação ao suporte de padrões de novas funcionalidades [Mohorovi i 2013]. A falta de suporte se agrava quando, além de considerarmos diferentes navegadores, consideramos também diferentes versões de cada navegador.
2. **Diferentes resoluções:** Existe uma grande variedade de tipos de dispositivos conectados à internet. Desktops, celulares, tablets, Smart TVs, variam de 3 polegadas à 50 polegadas ou mais [Perakakis et al. 2015].
3. **Depende de conexão:** Usuários trazem consigo seus dispositivos móveis a todo o momento, fazendo com que seja necessário o funcionamento das aplicações em qualquer lugar independente da qualidade ou existência de conexão [Desruell and Gielen 2015].

4. **Tempo e custo de carregamento das páginas:** Pesquisas indicam que mais de 50% dos visitantes desistem após esperarem por 4 segundos para o carregamento da página [Mohorovi i 2013]. O carregamento de códigos e imagens que não são utilizados nos websites em suas versões móveis faz com que requisições HTTP sejam realizadas desnecessariamente, aumentando o tempo de carregamento da página.
5. **Desktop-first**: Mais da metade dos acessos à websites vem de dispositivos móveis [Koetsier et al. 2005]. Entretanto a web ainda é desktop-first, onde boa parte dos websites são desenvolvidos com o objetivo de serem renderizados em monitores convencionais. Isto prejudica a visualização do conteúdo e a navegação através do *touch screen* em celulares e tablets. No caso das Smart TVs os problemas são em relação ao uso do controle remoto com fonte de entrada de dados. [Perakakis et al. 2015]. Em ambos os casos, a adoção de interfaces voltadas para tecnologias *touch-screen* resolvem ou, ao menos, amenizam o problema.

Em relação aos aplicativos nativos:

1. **Heterogeneidade de sistemas operacionais:** Cada sistema operacional de dispositivos móveis exige linguagens e ferramentas específicas para o desenvolvimento de aplicações nativas, fazendo com que um aplicativo escrito para uma plataforma não funcione em outra ([Malavolta et al. 2015] e [Xanthopoulos and Xinogalos 2013]).
2. **Tempo e custo de desenvolvimento:** O número de linguagens e ferramentas diferentes necessárias para o desenvolvimento de várias aplicações para funcionar em várias plataformas faz com que o custo de desenvolvimento seja grande [Espada et al. 2012] pois exige a contratação de profissionais experientes em diferentes sistemas. O tempo de desenvolvimento também é elevado por ser necessário o desenvolvimento de diferentes bases de código para cada plataforma, além de testes em diversos dispositivos.

QP2: Quais são as características esperadas de um *progressive web app*?

Não foram encontrados resultados nas bases de pesquisa que descrevessem, ou fizessem referência o conceito de *progressive web apps*. Possivelmente, isto se deve pelo fato desta metodologia ser recente e ainda não totalmente consolidada no mercado, tampouco na área acadêmica. Entretanto, a extração de resultados pelo buscador *Google* retornou diversas publicações que definem clara e diretamente o conceito de *progressive web apps*. Este conceito foi definido por desenvolvedores da Google em [Google 2016] como sendo "experiências que combinam o melhor da web e o melhor das aplicações nativas". Também definem 10 pontos específicos que caracterizam um *progressive web apps*:

1. Progressivo - Funciona para todos os usuários, independente e dispositivos e da escolha de navegador pois é construída utilizando os conceitos de *progressive enhancement*.
2. Responsivo - Se adapta às resoluções de qualquer dispositivo: desktop, celulares, tablets, TVs, entre outros.
3. Independente de conexão à Internet - Utiliza *service workers* para funcionar offline ou em conexões de baixa qualidade.
4. *App-like* - Experiência de aplicativo nativo de navegação e interação porque é construída sobre o modelo de *app-shell*.

5. Atualizado - Também utiliza *service workers* para manter seu conteúdo sempre atualizado.
6. Seguro - Servida via HTTPS para prevenir *snooping* e assegurar que o conteúdo não foi alterado.
7. Detectável - É identificada como aplicação, graças ao manifesto da W3C e também ao uso de *service workers*, permitindo também ser encontrada por motores de buscas.
8. Melhor engajamento - Possibilita o envio de notificações push para melhorar o engajamento do usuário com a aplicação.
9. Instalável - Permite que usuários mantenham aplicações na tela inicial dos seus dispositivos sem o intermédio de uma loja de aplicativos.
10. Passível de ligação via URL (*linkable*) - Facilmente compartilhável via URL. Não necessita instalação complexa.

Apesar de o conceito de *progressive web apps* não aparecer nos resultados das bases de pesquisa, diversos artigos extraídos provêm embasamento científico para alguns dos pontos que definem um *progressive web app*.

Em [Desruell and Gielen 2015] *progressive enhancement* é definido como uma abordagem que visa maximizar a experiência de usuário em diferentes ambientes. Um desenvolvimento progressivo seguindo este conceito força o desenvolvedor a considerar ambientes com suporte a um número menor de funcionalidades desde o início do processo de desenvolvimento, fazendo com que a aplicação entregue boa experiência independente da qualidade do ambiente. [Wells and Draganova 2007] divide o processo de desenvolvimento de aplicações com *progressive enhancement* em três camadas: conteúdo, apresentação e comportamento. O conteúdo é descrito utilizando apenas uma linguagem de marcação, interpretada por qualquer navegador independente de CSS e javascript. A apresentação consiste em dois arquivos de estilo, sendo um chamado básico, contendo estilos suportados por navegadores mais antigos, e um chamado moderno, contendo estilos mais aprimorados, para navegadores modernos. A última camada contém scripts que definem o comportamento da página, desenvolvido utilizando ferramentas que garantem a compatibilidade entre navegadores. Em [Mohorovi i 2013] o termo *progressive enhancement* é utilizado em relação ao desenvolvimento de interfaces na metodologia *mobile-first*, nas quais o *layout* é desenvolvido primeiramente considerando as resoluções menores e progressivamente aprimorado à medida em que a resolução aumenta. Além de descrever *progressive enhancement,* [Hall 2009] também fala sobre o seu oposto, *graceful degradation*. Ele mostra que ao invés de desenvolver para os ambientes mais simples primeiro como na abordagem progressiva, nessa abordagem a aplicação é construída primeiramente com o foco nos ambientes mais modernos. Caso esta seja executada em um ambiente inferior, funcionalidades que utilizam recursos não disponíveis são bloqueados, fazendo com o que o usuário não tenha uma experiência completa.

Diversos autores descreveram o conceito de web design responsivo como sendo uma abordagem em que ao invés de desenvolver diversas versões de um website para diversos dispositivos diferentes, cria-se um único *layout* capaz de adaptar-se à resolução do dispositivo em que se encontra. Faz uso de *grids* fluídos, vídeos e imagens de tamanhos flexíveis e *media queries*, que permitem a execução condicional de scripts de estilo em relação a resolução e/ou orientação da tela ([Walsh et al. 2015], [Majid et al. 2015], [Johansen et al. 2013], [Fabri et al. 2013] e [Mohorovi i 2013], [Yang and Mei 2014], [Lee et al. 2015], [Jiang et al. 2014], [Mullins 2015], [Hernández-Nieto 2015], [Katajisto 2015]).

[Tahir et al. 2015] descreve o conceito de *web workers* como uma forma de utilizar aplicações web *offline*. Esta é uma API que define uma forma de executar scripts

em background em paralelo aos scripts sendo executados na página. [Russel 2015] descreve *service workers* como sendo um tipo de *web worker*. Este é o script executado em paralelo que torna possível que sejam executados de rotinas periodicamente para a sincronização de dados armazenados localmente, para que a aplicação esteja sempre atualizada.

QP3: Quais problemas atuais da web e de aplicações nativas podem ser resolvidos com *progressive web apps*?

Comparando as respostas das questões QP1 e QP2, podemos constatar que todos os pontos problemáticos no desenvolvimento de aplicações web e nativas podem ser totalmente resolvidos com o desenvolvimento de *progressive web apps*. É necessário constatar que existem diversos outros pontos desafiadores para as plataformas web e móvel, porém neste trabalho estamos considerando apenas os pontos apontados pelos autores dos artigos selecionados.

As características de progressividade e responsividade dos *progressive web apps* fazem com que o seu acesso possa ser realizado de qualquer dispositivo, independente da resolução ou sistema operacional. A independência de conexão à internet provê uma melhora na experiência de usuário quando esse se encontra em locais sem conexão, ou durante uma queda de conexão. Ao invés de redirecioná-lo à uma página de erro, esse estado será tratado conforme a necessidade de conexão, possibilitando o usuário a continuar utilizando a aplicação com os conteúdos armazenados previamente em *cache*. A utilização do modelo *app-shell* permite que a aplicação armazene o código para a renderização da estrutura básica da aplicação (i.e. menus) tornando assim o carregamento inicial muito mais rápido. Além de também prover uma melhor experiência de usuário do que uma aplicação web comum por ter comportamento similar à aplicativos nativos [Google 2016].

3. Considerações finais

O objetivo deste trabalho foi realizar uma revisão sistemática sobre aplicações web e aplicações móveis utilizando tecnologias web com o foco no conceito de *progressive web apps*. Foi constatado que ainda não há na literatura acadêmica referências sobre esta metodologia. Mesmo assim, a realização dessa revisão permitiu com que fossem levantadas referências bibliográficas que validam as funcionalidades propostas por este conceito. Diversos pontos negativos sobre as plataformas web e móvel foram citados pelos trabalhos extraídos, bem como algumas de suas soluções. Ao fim concluímos que o desenvolvimento de um *progressive web app* resulta em um aplicação que pode resolver muitos dos problemas atuais, fazendo uso de recursos modernos do ambiente móvel, porém se adaptando também em navegadores antigos.

Como trabalho futuro, propomos um estudo sobre outros métodos para o desenvolvimento de aplicações móveis através de tecnologias para web, além da comparação desses quanto ao custo de desenvolvimento e qualidade da experiência do usuário em um caso real de aplicação dentro do contexto nacional brasileiro, que por sua vez tornam agudas as restrições de contexto.

Referências

[Kitchenham 2004] Kitchenham, B. (2004). “Procedures for Performing Systematic Reviews. Joint Technical Report”, Keele, UK, Keele University, 33(2004):1-26.

[Gibbs 2016] Gibbs, S. (2016). “Mobile web browsing overtakes desktop for the first time”. https://www.theguardian.com/technology/2016/nov/02/mobile-web-browsing-desktop-smartphones-tablets (Acessado em 13 de Novembro de 2016)

[Koetsier 2005] Koetsier, J. (2005). “Wait, what? Mobile browser traffic is 2X bigger than app traffic, and growing faster”. http://venturebeat.com/2015/09/25/wait-what-mobile-browser-traffic-is-2x-bigger-than-app-traffic-and-growing-faster/ (Acessado em 12 de Outubro de 2016)

[Google 2016b] Google, Inc. (2016). “Your First Progressive Web App”. https://developers.google.com/web/fundamentals/getting-started/codelabs/your-first-pwapp/ (Acessado em 12 de Outubro de 2016)

[Russel 2015] Russel, A, et al. (2015) “Service Workers W3C Working Draf”. https://www.w3.org/TR/service-workers/ (Acessado em 12 de Outubro de 2016)

[Xanthopoulos and Xinogalos 2013] Xanthopoulos, S. and Xinogalos, S. (2013). “A comparative analysis of cross-platform development approaches for mobile applications”, Proceedings of the 6th Balkan Conference in Informatics

[Majid et al. 2015] Majid, E.; Kamaruddin, N.; Mansor, Z. (2015). “Adaptation of usability principles in responsive web design technique for e-commerce development”, International Conference on Electrical Engineering and Informatics (ICEEI).

[Perakakis et al. 2015] Perakakis E.; Ghinea, G.; Thanou, E. (2015), “Are Websites Optimized for Mobile Devices and Smart TVs?”, 8th International Conference on Human System Interaction (HSI).

[Walsh et al. 2015] Walsh, T.; McMinn, P.; Kapfhammer, G. M. (2015) “Automatic Detection of Potential Layout Faults Following Changes to Responsive Web Pages”, 30th IEEE/ACM International Conference on Automated Software Engineering (ASE).

[Desruell and Gielen 2015] Desruell, H. and Gielen, F. (2015) “Context-Driven Progressive Enhancement of Mobile Web Applications: A Multicriteria Decision-Making Approach”, The Computer Journal, vol. 58(8)

[Johansen et al. 2013] Johansen, R; Britto, T. C. P.; Cusin, C. A. (2013) “CSS browser selector plus: a JavaScript library to support cross-browser responsive design”, WWW 2013 Companion Proceedings of the 22nd International Conference on World Wide Web

[Fabri et al. 2013] Fabri, D.; Krempser, T.; Filgueiras, L. V. L. (2013) “Estudo de Responsive Web Design Aplicado a um Sistema de Pesquisa de Opinião na Área Médica”, IHC '13 Proceedings of the 12th Brazilian Symposium on Human Factors

in Computing Systems

[Espada et al. 2012] Espada, J.; Crespo, R. G.; Martínez, O. S.; G-Bustelo, B. C. P.; Lovelle J. M. C. (2012) "Extensible architecture for context-aware mobile web applications", Expert Systems with Applications Volume 39, Issue 10, August 2012

[Mohorovi i 2013] Mohorovi i , S. (2013), Implementing Responsive Web Design for Enhanced Web Presence, 36th International Convention on Information & Communication Technology Electronics & Microelectronics (MIPRO)

[Wells and Draganova 2007] Wells, J. and Draganova, C. (2007), "Progressive Enhancement in the Real World", HT '07: Proceedings of the eighteenth conference on Hypertext and hypermedia Pages 55-56

[Tahir et al. 2015] Tahir, Z.; Inamoto, T.; Higami, Y.; Kobayashi, S. (2015) "The analysis of Automated HTML5 Offline Services (AHOS)", 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)

[Malavolta et al. 2015] Malavolta, I.; Ruberto, S.; Soru, T.; Terragni, V. (2015). "End Users' Perception of Hybrid Mobile Apps in the Google Play Store" International Conference on Mobile Services (MS), 2015 IEEE

[Yang and Mei 2014] Yang, H. and Mei, H. (2014) "Responsive Universal Design with Universal User Profiles and CSS User Queries" Fourth International Symposium on Business Modeling and Software Design

[Lee et al. 2015] Lee, J.; Lee, I.; Kwon, I.; Yun, H.; Lee, J.; Jung, M.; Kim, H. (2015) "Responsive Web Design According to the Resolution". 2015 8th International Conference on u- and e-Service, Science and Technology

[Jiang et al. 2014] Jiang, W.; Zhang, M.; Zhou, B.; Jiang, Y.; Zhang Y. (2014) "Responsive Web Design Mode and Application". 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)

[Mullins 2015] Mullins C. (2015) "Responsive, Mobile App, Mobile First: Untangling the UX Design Web in Practical Experience" SIGDOC '15 Proceedings of the 33rd Annual International Conference on the Design of Communication Article No. 22

[Hernández-Nieto 2015] Hernández-Nieto, C.; Sánchez, X.; Salinas, P. (2015) "A Mobile First Approach for the Development of a Sustainability Game". 2015 International Conference on Virtual and Augmented Reality in Education

[Hall 2009] Hall, C. A. (2009) "Web Presentation Layer Bootstrapping for Accessibility and Performance" Proceedings of the International Cross-Disciplinary Conference on Web Accessibility, W4A 2009, Madrid, Spain, April 20-21, 2009

[Katajisto 2015] Katajisto, L. (2015) “Creating Support Content for Responsive Websites at Microsoft Mobile” 2015 IEEE International Professional Communication Conference (IPCC)

Proposta de Melhoria do Processo de Produção com a Implantação de Integração Contínua em uma Empresa de Desenvolvimento de Software

Cleiviane Cardoso da Costa

Escola Superior Aberta do Brasil (ESAB)
Avenida Santa Leopoldina, 840, 29102-040 – Vila Velha – ES – Brasil.

{costa.cleiviane@gmail.com}

Abstract. *During the software life cycle is natural that changes happen. For that changes be integrated into existing software you need to perform an activity known as deployment. Conduct the process of deployment in software development companies tends to be a manual activity, labor intensive and susceptible to human errors. Due to their repetitive nature, it is clear that this process can be automated. Continuous Integration is a practice that proposes to streamline and automate the deployment and hence reduce the occurrence of human errors, integrating the changes more often. The purpose of this study was to guide the implementation of Continuous Integration in the company Tron Informática.*

Resumo. *Durante o ciclo de vida do software é natural que mudanças aconteçam. Para que as mudanças sejam integradas ao software existente é preciso realizar uma atividade conhecida como deploy. Conduzir o processo de deploy nas empresas de desenvolvimento de software tende a ser uma atividade manual, trabalhosa e suscetível a erros humanos. Devido a sua natureza repetitiva, é evidente que este processo pode ser automatizado. A Integração Contínua é uma prática que propõe agilizar e automatizar o deploy e, consequentemente, diminuir a ocorrência de erros humanos, integrando as mudanças frequentemente. A proposta deste trabalho foi orientar a implantação de Integração Contínua na empresa Tron Informática.*

1. Introdução

A importância dos programas de computadores para os ambientes empresariais vem crescendo na medida em que o mercado consumidor torna-se diariamente mais exigente no que tange a qualidade e a produtividade dos fornecedores. Atendendo a esta demanda, processos que antes eram executados manualmente, passam a ser realizados por computadores em velocidades cada vez maiores. Entretanto, as necessidades do mercado e dos usuários são mutáveis e os programas de computador precisam responder a estas mudanças, também de forma cada vez mais rápida [Lopes, 2005].

Uma mudança pode ser solicitada para atender um requisito legal, uma nova necessidade do usuário ou uma atualização de tecnologia, porém, se mal gerenciadas as mudanças podem causar confusão entre os membros do time que está trabalhando no projeto [PRESSMAN, 2010]. Esta situação é ainda mais evidente em equipes que

trabalham sob metodologias tradicionais, uma vez que, estas metodologias possuem uma estrutura sequencial que não prevê alterações de escopo no andamento do projeto.

Afim de tornar possível construir o *software* com base em um escopo dinâmico, que permita atender as necessidades de seus clientes, muitas empresas de desenvolvimento de *software* adotaram a Metodologia Ágil em seus projetos para, entre outras coisas, trazer fluidez ao processo de mudança no *software*. Este foi o caso da Tron Informática Ltda., empresa que desenvolve *softwares* para o segmento contábil desde 1992.

A Tron Informática já utiliza algumas práticas de metodologia ágil como a adoção do *Scrum* em seus projetos onde desenvolve produtos web. Agora a empresa almeja dar mais um passo no sentido de adotar a agilidade como norteadora, melhorando seu processo de produção, o momento em que uma funcionalidade será disponibilizada para uso.

Diante deste cenário, o objetivo desta pesquisa é analisar o processo de produção dos produtos web da empresa Tron Informática Ltda. e implantar a Integração Contínua, visando automatizar este processo afim de diminuir o tempo gasto sem deixar de garantir a alta disponibilidade e a qualidade destes produtos.

O artigo está estruturado da seguinte forma: na seção 2 é feita a fundamentação teórica; na seção 3 são descritas as etapas realizadas para alcançar o objetivo deste trabalho bem como a análise dos resultados obtidos e, por fim, na seção 4 está a conclusão e as perspectivas de trabalhos futuros.

A natureza deste trabalho é um estudo de caso. Gerhardt (2009) afirma que o estudo de caso é um estudo de uma entidade bem definida como um programa, uma instituição, um sistema educativo, uma pessoa ou uma unidade social. Do ponto de vista dos objetivos, trata-se de uma pesquisa exploratória, onde busca-se coletar informações e entender da melhor forma possível todo o processo que envolve a empresa objeto deste estudo de caso. Do ponto de vista da abordagem, trata-se de uma pesquisa qualitativa, pois não se tem a preocupação com a representatividade numérica e sim com a compreensão e melhoria do processo de produção da organização.

2. Fundamentação Teórica

2.1 Metodologia Ágil

Pressman (2010) afirma que o *software*, assim como todo sistema complexo, evolui ao longo do tempo. Para que estas evoluções aconteçam o *software* passa por diversas mudanças sejam estas para cumprir um requisito legal, uma adequação tecnológica ou para atender as necessidades do negócio ou do usuário. Mudanças no *software* são caras, principalmente se forem feitas sem controle e mal gerenciadas [PRESSMAN, 2010].

Em 2001, Kent Beck e outros dezesseis renomados desenvolvedores, autores e consultores da área de *software* assinaram o "Manifesto para o Desenvolvimento Ágil de *Software*", propondo uma adequação nos processos das empresas de desenvolvimento de *software*, para que elas pudessem responder as mudanças de forma mais rápida e apropriada [PRESSMAN, 2010]. Assim popularizou-se o termo Metodologia Ágil ou Método Ágil.

O Manifesto Ágil baseia-se em quatro valores fundamentais [FOWLER, 2006]:

1. Indivíduos e interações, ao invés de processos e ferramentas.
2. *Software* executável, ao invés de documentação.
3. Colaboração do cliente, ao invés da negociação de contratos.
4. Respostas rápidas à mudança, ao invés de seguir planos.

De acordo com Fagundes (2005), as filosofias contidas no Manifesto Ágil foram refinadas por seus criadores em uma coleção de doze princípios, para facilitar seu entendimento. Estes princípios, que devem ser seguidos por todos os métodos ágeis, são:

1. A prioridade é satisfazer ao cliente através de entregas de *software* de valor contínuas e frequentes;
2. Receber bem as mudanças de requisitos, mesmo em uma fase avançada, dando aos clientes vantagens competitivas;
3. Entregar *softwares* em funcionamento com frequência de algumas semanas ou meses, sempre na menor escala de tempo;
4. As equipes de negócio e de desenvolvimento devem trabalhar juntas diariamente durante todo o projeto;
5. Manter uma equipe motivada fornecendo ambiente, apoio e confiança necessário para a realização do trabalho;
6. A maneira mais eficiente da informação circular dentro da equipe é através de uma conversa face a face;
7. Ter o *software* funcionando é a melhor medida de progresso;
8. Processos ágeis promovem o desenvolvimento sustentável. Todos envolvidos devem ser capazes de manter um ritmo de desenvolvimento constante;
9. Atenção contínua a excelência técnica e um bom projeto aumentam a agilidade;
10. Simplicidade é essencial;
11. As melhores arquiteturas, requisitos e projetos provêm de equipes organizadas;
12. Em intervalos regulares, a equipe deve refletir sobre como se tornarem mais eficazes e então se ajustar e adaptar seu comportamento.

2.2 Controle de Versão

Um sistema de controle de versão (ou *Source Code Management* ou SCM) é muito importante para que várias pessoas possam trabalhar em um mesmo projeto, pois mantém todas as diferentes versões de um *software* centralizadas em um único local. É sua função identificar, organizar, registrar e controlar todas as modificações realizadas por cada membro da equipe. Seu objetivo é maximizar a produtividade minimizando erros [PRESSMAN, 2010].

Geralmente o SCM mantém os artefatos de configuração (que podem ser o código fonte, documentação, bases de dados e arquivos de configuração) em um local

centralizado denominado repositório. Quando um membro é adicionado a equipe ele clona uma versão estável desse repositório (a *baseline*) em seu ambiente local para fazer suas alterações. As novas alterações são enviadas novamente para o repositório, que deve atualizar a *baseline* e manter um registro com os dados de quem fez a alteração bem como da versão antes e depois da alteração, caso seja necessário voltar o sistema para um ponto que estava estável sem alterar a *baseline*.

Outro ponto bastante importante é o uso de braços (ou *branchs*), que são ramificações geradas a partir da *baseline*. Criar um *branch* significa dizer que você vai divergir da linha principal de desenvolvimento e continuar a trabalhar em outra linha sem bagunçar a linha principal [CHACON; STRAUB, 2014]. Isso mantém o repositório organizado e faz uma separação entre o que está em desenvolvimento, o que está em ambiente de teste e o que está em ambiente real ou ambiente de produção.

2.3 Integração de Software

Em Desenvolvimento de sistemas, Integração de *Software* é processo em que um conjunto de funcionalidades (versão) são integradas ao *software* que já está em execução e acessível aos seus usuários finais. O processo de produção, também conhecido como *deploy*, é o momento em que o *software* passará por uma integração.

Para que o *deploy* seja feito, um *build* é gerado com a nova versão e enviado para o servidor onde o *software* está em funcionamento. Duval (2007) explica que um *build* é mais que a compilação de um *software* (ou as variações dinâmicas da linguagem), e pode consistir também em teste, inspeção, documentação e outras ações. Um *build* atua como o processo de unir todo o código fonte e seus artefatos e verificar se o *software* funciona da forma correta como uma unidade coesa [DUVALL, 2007].

Geralmente, para realizar um *deploy* são executados alguns passos: atualizar o código fonte usando o sistema de controle de versões; executar os testes unitários e verificar que todos estão passando; instalar novos *softwares*, *plugins* e dependências que podem ter sido adicionados; criar pastas, copiar arquivos, alterar permissões, etc.; executar scripts com alteração em banco de dados; reiniciar o *software* e monitorá-lo para garantir que tudo está funcionando como esperado.

2.4 Integração Contínua

Integração Contínua (ou *Continuous Integration* ou CI) é uma prática adotada por times de desenvolvimento ágil onde cada membro integra seu trabalho frequentemente, pelo menos uma vez ao dia, o que leva a múltiplas integrações diariamente [FOWLER,2006]. Sabendo disso, utilizamos esta prática para trazer eficiência ao processo de produção nas empresas de desenvolvimento de *software*, propondo que erros sejam identificados e corrigido o quanto antes.

A cada integração o *software* passa por validações automatizadas para detectar erros de integração tão rápido quanto possível. As integrações são feitas em um repositório central, denominado Servidor de Integração Contínua, onde é instalada uma ferramenta que auxilia todo o processo de CI.

Smart (2011) explica que, em essência, Integração Contínua é sobre reduzir riscos causados por integrações através um *feedback* rápido. A fim de tornar isso

possível, as ferramentas de CI monitoram o sistema de controle de versões procurando por novas atualizações. Quando um desenvolvedor envia uma nova alteração para o controle de versão, esta ferramenta executa automaticamente os testes unitários para identificar a presença de falha. Se uma falha é encontrada, a ferramenta notifica imediatamente os desenvolvedores para que eles possam, rapidamente, fazer as correções necessárias [SMART, 2011]. Se nenhuma falha for detectada, a ferramenta gera um *build* incluindo as alterações em uma versão do *software*.

Fowler (2006) explica que o ciclo da Integração Contínua envolve as seguintes etapas:

1. Fazer uma cópia local do repositório (conhecida como clone);
2. Fazer as alterações necessárias no código, usando teste unitário;
3. Gerar um *build* do projeto, sem nenhum teste falhando;
4. Enviar as alterações para o repositório remoto;
5. Atualizar o código local incluindo as alterações de outros desenvolvedores;
6. Gerar novamente um *build*, fazendo as correções para garantir que nenhum teste esteja falhando;
7. Enviar novamente as alterações para o repositório remoto;
8. A Ferramenta de CI irá identificar a nova versão do código e executar todas as etapas para gerar seu próprio *build* e atualizar o *software*.

O ciclo é finalizado quando as mudanças que foram envidadas para o repositório passam pelo *build* da Integração Contínua com sucesso. A figura 1, ilustra o ciclo da CI e mostra que os programadores enviam suas alterações para o Sistema de Controle de Versão (CVS), em contrapartida, o Servidor de Integração Contínua (CIS) está monitorando o CVS e executando um *build* em cada alteração enviada. Cada *build* é concluído com um status: sucesso ou falha.

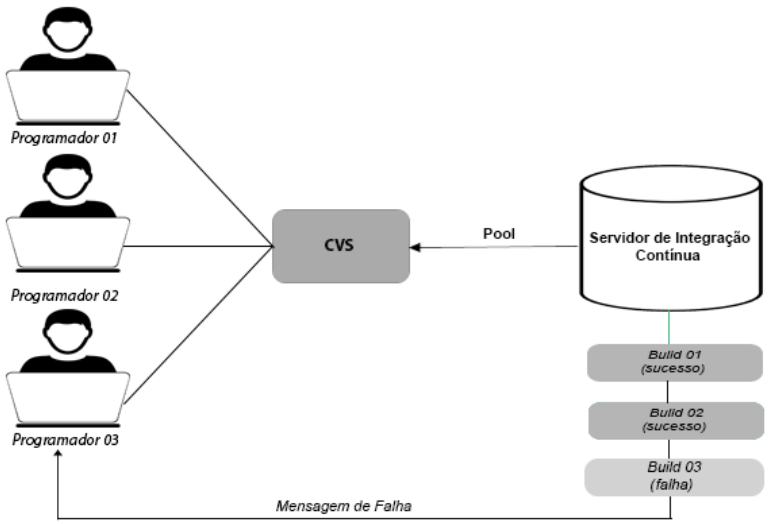


Figura 1 – Ciclo de funcionamento da CI

2.5 Principais Ferramentas para Integração Contínua

Existem várias ferramentas para integração contínua disponíveis no mercado, a maioria já atingiu um alto nível de maturidade atingindo com excelência o proposito para o qual foi destinada. Destacamos a seguir, algumas das principais ferramentas usadas para CI.

Team City

TeamCity é uma poderosa ferramenta de Integração Contínua, cuja principal característica é a execução de testes automáticos [JETBRAINS, n.d.]. Bastante utilizada para projetos Java e .Net, possui duas versões: uma com licença comercial (*Enterprise Edition*) e uma não comercial (*Professional Edition*). A versão *Professional* permite que sejam configurados 20 *builds* e que 3 agentes sejam utilizados. Já a versão *Enterprise* permite um número ilimitado de *builds* e o número de agentes vai de acordo com o pacote adquirido.

O principal diferencial do TeamCity é sua interface amigável que facilita a usabilidade. Outros recursos que merecem destaque são a criação de templates de configurações; construção, verificação e execução de testes automatizados; mais de 100 *plugins* para adicionar novas funcionalidades; arquitetura escalável e relatório de progresso *on the fly,* onde é possível acompanhar os possíveis erros antes mesmo do *build* finalizar.

Jenkins

Jenkins é o principal servidor de integração contínua de código aberto. Construído em Java, seu objetivo é aumentar a produtividade, oferecendo centenas de *plugins* para apoiar a construção, os testes, a implantação e automatização de projetos de *software*. Todos estes processos são feitos através do Jenkins de forma contínua tornando mais fácil para os desenvolvedores integrar alterações ao projeto, e aos usuários usufruírem de uma nova versão [JENKINS, n.d.].

Os maiores destaques do Jenkins estão na fácil instalação e configuração; rico ecossistema de *plugins*; fácil customização; e facilidade para carregar os *builds* e testes entre computadores com sistemas operacionais diferentes. Isto é útil, por exemplo, quando um produto possui versão para Windows e Linux ou quando uma empresa desenvolve diferentes projetos que rodam em diferentes sistemas operacionais.

CruiseControl

CruiseControl é tanto uma ferramenta quanto um *framework* (conjunto de ferramentas) extensível que auxilia na construção de um processo personalizado de Integração Contínua [CRUISECONTROL, n.d.]. Isto inclui um conjunto de dezenas *plugins* para controle de código fonte, compilação de *software* e notificações por e-mail. Através de uma interface web, o usuário obtém informações sobre a versão atual e as passadas. A ferramenta é *open source* e escrita em Java, possuindo integração com uma ampla variedade de projetos como Ant, Maven, Rake, etc.

3. Apresentação dos Resultados

3.1 Cenário Atual

Parte fundamental para a implantação de um processo de CI é o conhecimento do processo de produção atual. Destarte, as etapas executas pela Tron Informática para liberação de uma versão são: o programador realiza uma correção ou desenvolve uma nova funcionalidade; o responsável pela Integração (integrador) realiza *deploy* no servidor de testes; a equipe de testes homologa a versão; e então, o integrador realiza *deploy* da versão homologada no servidor de produção. Todo este processo é executado manualmente, o tempo gasto com este processo é grande e o crescimento exponencial na medida em que o número de projetos da empresa cresce.

Outro ponto fundamental é que as aplicações desenvolvidas pela empresa precisam de um ambiente heterogêneo para serem executadas. Por este motivo, a ferramenta de Integração Contínua adotada deve ser capaz de executar todo seu processo independente de sistema operacional, banco de dados, linguagens de programação ou demais configurações específicas de tais ambientes.

3.2 Escolha da Ferramenta

DUVALL (2007) explica que para escolher uma ferramenta de CI é necessário considerar o ambiente e o processo de desenvolvimento e que alguns fatores devem influenciar na decisão como: as funcionalidades, compatibilidade com o ambiente e outras tecnologias utilizadas, nível de confiabilidade, quantidade de usuários e mantenedores (que garantirá longevidade) e o grau de usabilidade. Considerando estes fatores e conhecendo o cenário atual da Tron Informática, descrito na subseção 3.1, a ferramenta escolhida foi a Jenkins.

3.3 Implantação

Para a realização deste trabalho, definiu-se que a implantação inicial do processo de integração contínua na empresa contemplaria três produtos web: um produto que roda em ambiente Windows e dois produtos que rodam em ambiente Linux.

Antes de iniciar a configuração do Jenkins, é necessário realizar a seguintes pré configurações no servidor onde a ferramenta de CI será configurada: instalar versão recente do Java Development Kit (JDK), possuir um Sistema de Controle de Versão operando corretamente e instalar a ferramenta Jenkins.

Após concluídas as instalações pré requisitadas é necessário adicionar os *plugins* básicos para o funcionamento da CI. Essa adição é feita na opção *Jenkins → Manage Jenkins → Manage Plugins*. Os *plugins* necessários para este trabalho foram:

- Node and Label Parameter Plugin, que auxiliará a criação do ambiente *master*/slave;
- SSH Plugin, utilizado pelo Jenkins para conectar em outras máquinas;
- Git Plugin, possibilitará que o Jenkins monitore o sistema de controle de versão;
- Mailler Plugin, que fornece ao Jenkins a funcionalidade de enviar e-mails informando sobre o status das integrações.

Outros *plugins* podem ser necessários dependendo da configuração do servidor onde o Jenkins for instalado ou das necessidades do projeto.

3.4 Criação do Ambiente *Master/Slave*

Para dar suporte ao cenário mostrado na subseção 3.1, o Jenkins deve funcionar sobre uma arquitetura *master/slaves* (ou mestre/escravos), onde as tarefas (*job's*) são criadas no nó *master* e executados, de fato, nos nós *slaves*. A função do *master*, portanto, é de gerenciar os nós e garantir que os *job's* serão executados no local adequado. Ao final da integração, cada *slave* deve informar ao *master* o resultado do trabalho executado.

Nesta arquitetura é possível configurar múltiplos nós como máquinas *slaves*, onde cada uma irá monitorar o ambiente de uma máquina real [JENKINS, n.d.]. A criação de nós é feita na opção *Jenkins* → *Manage Jenkins* → *Manage Nodes*, onde são adicionados nós com as informações sobre cada máquina slave.

3.5 Criação dos *Jobs*

SMART (2011) descreve um *job* como um passo ou tarefa específica a ser executado no processo de *build*. Um *job* pode incluir apenas compilar o código fonte e executar os testes unitários, ou pode envolver outras tarefas relacionadas como executar testes de integração, medir cobertura de código, gerar documentação técnica, fazer *deploy* da aplicação para um servidor web, entre outras.

Conforme apresentado na figura 2, para criar um *job* no Jenkins é preciso informar o nome e o estilo do projeto.

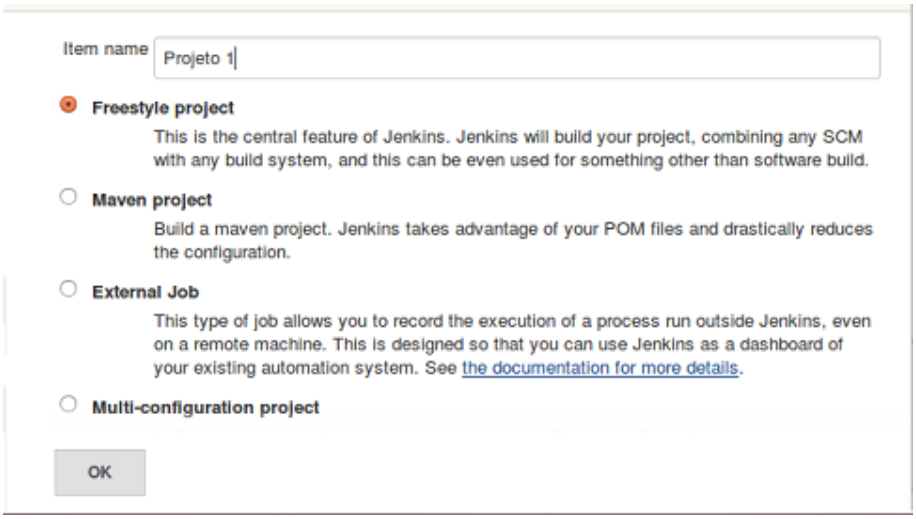


Figura 2. Criação de um *job*

Para associar o *job* criado ao *slave* correspondente é preciso marcar a opção "Este *build* é parametrizado" e selecionar o nó onde foi configurado o *slave* desejado, conforme exibido na figura 3.

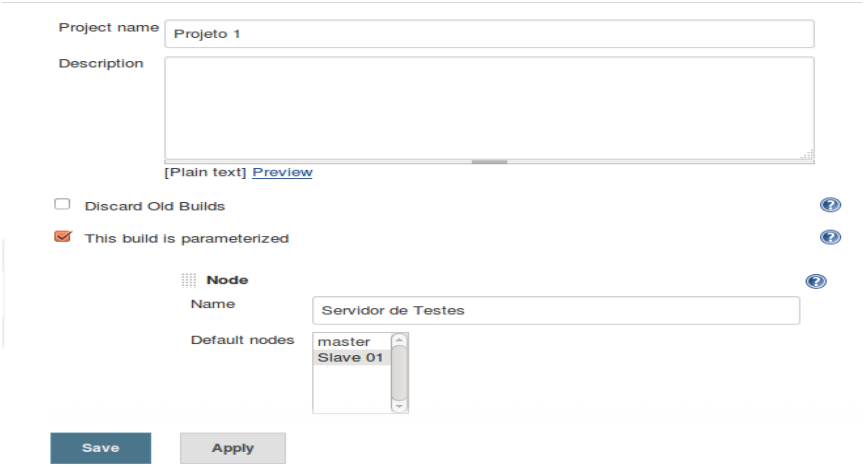


Figura 3 - Associando o *build* ao *slave*

Por último, deve-se informar as demais configurações do *build* como definições sobre ferramenta de controle de versão, repositório de código fonte, *branch* utilizado, em que momento o *build* deve ser executado, configurações de construção do *build* e ações pré e pós *build*. Todas as etapas do *build* também devem ser adicionadas, e, normalmente, estão associadas as funcionalidades providas pelos *plugins* instalados ou também à execução de testes e *feedback* providos pelo próprio Jenkins.

3.6 Execução de um *Job*

O Jenkins irá realizar a execução dos job's obedecendo o intervalo de tempo definido na configuração de cada *job* ou, manualmente, caso o usuário opte por esta opção. Para que o *job* configurado na subseção 3.5 execute usando o *slave* criado na subseção 3.4, é preciso utilizar a opção "*Build with Parameters*" e selecionar o *slave* desejado, conforme mostra a figura 4.

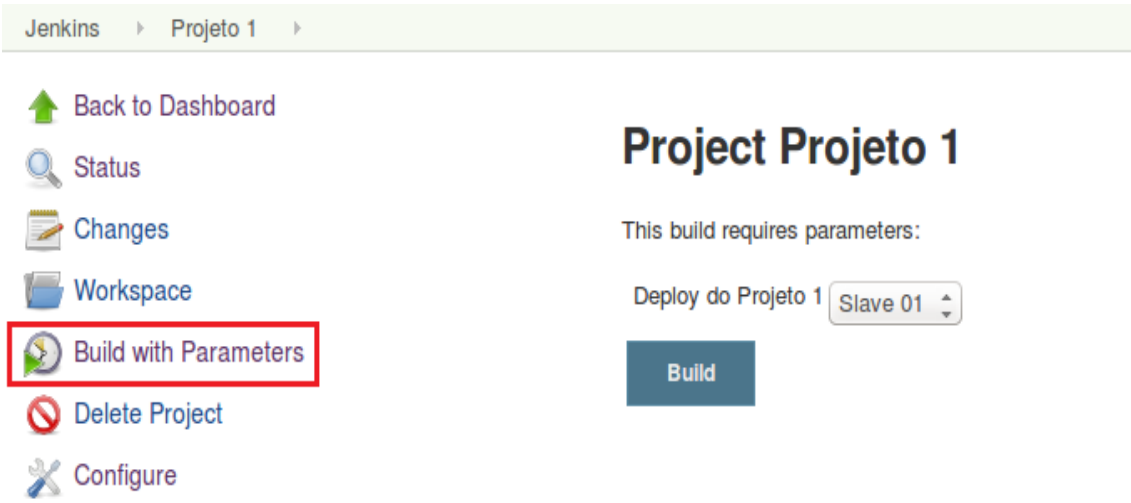


Figura 4 - Executando o *build* do projeto 1

3.7 Automatização dos *Jobs*

Um dos objetivos deste trabalho é fazer com que o processo de produção se torne mais rápido e eficiente e, para tal, ele precisa ser o mais automatizado possível. O Jenkins possui a funcionalidade de iniciar a execução dos *build's* de forma automática. Ao configurar a opção *"Build Triggers"*, é informado ao Jenkins quando ele irá disparar o *job* e executar os *build's* [SMART, 2011]. Os *triggers* podem ser: iniciar o *job* quando outro *job* for finalizado; iniciar o *job* em intervalos definidos; ou iniciar o *job* quando uma houver mudanças no SCM.

A figura 5 mostra que para os projetos configurados neste trabalho, selecionou-se a opção para que o *build* seja executado sempre que houver uma nova atualização no repositório de código fonte (Poll SCM).

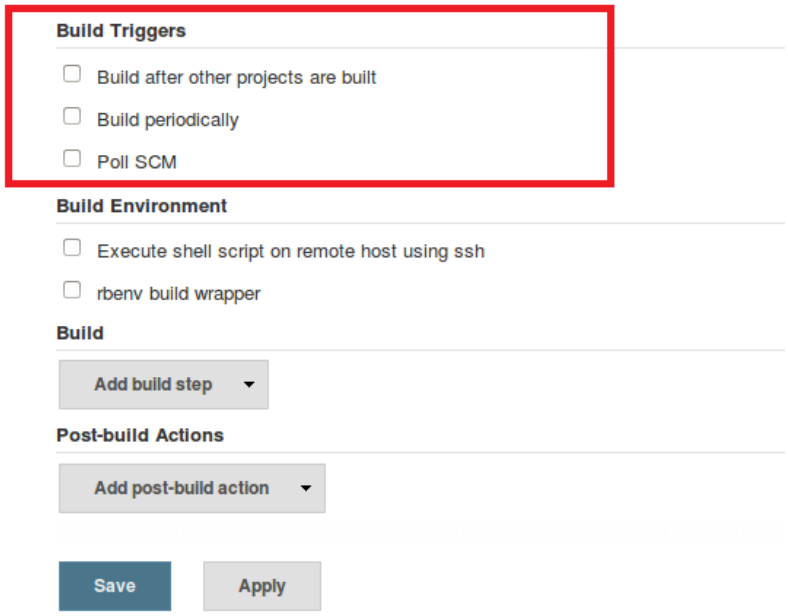


Figura 5 – Configuração do disparador de *build*

A tabela 1 apresenta as configurações dos três projetos adotados, bem como o resultado obtido com suas execuções.

Tabela 1 – Configuração e Resultados da CI nos Projetos implantados

Projeto	Sistema Operacional	Controle de Versão	Teste unitário	Intervalo de execução	Resultado do *build*
Projeto 01	Linux	GIT	✓	Poll SCM	Sucesso
Projeto 02	Linux	GIT	✓	Poll SCM	Sucesso
Projeto 03	Windows	GIT	✓	Poll SCM	Sucesso

3.8 Análise dos Resultados

FOWLER (2006) aponta que um dos objetivos da Integração Contínua é permitir que o time de desenvolvedores entregue *software* consistente mais rapidamente. Sabendo disso podemos afirmar que a adoção da Integração Contínua trouxe uma evolução no uso de metodologias ágeis para a empresa objeto de estudo deste trabalho, além da melhoria no paradigma de desenvolvimento absorvida pela equipe dos projetos implantados. Observamos que esta adoção trouxe outras melhorias para empresa tais como: diminuição do tempo gasto para fazer *deploy*; automatização de todo o processo de produção, que aumentou a segurança diminuindo os riscos de erro humano; possibilidade de pessoas alheias aos projetos terem acesso aos artefatos gerados.

A tabela 2 ilustra estas melhorias fazendo uma comparação entre as atividades antes e depois da implantação da Integração Continua (CI). Vale ressaltar que o campo "Tempo gasto com *deploy*" é uma soma entre o tempo gasto nos três projetos implantados.

Tabela 2: Atividades avaliadas com a implantação da CI.

Item da avaliação	Antes da CI	Depois da CI
Tempo gasto com *deploy*	85 min	12,41 min
Segurança contra falha humana.		✓
Execução automatizada de Testes	x	✓
Criação automatizada de artefatos da Integração	x	✓
Resultados das Integrações Transparentes a membros da equipe e externos a equipe	x	✓
Detecção das falhas ocorridas na execução do processo de integração	x	✓
Relatório indicando o resultado das integrações	x	✓
Relatório para visualizar os artefatos gerados	x	✓
Relatório indicando motivo de falhas durante as integrações	x	✓
Emissão automática de alerta contra falhas durante a execução das integrações	x	Envio de email
Exibição de qual a porcentagem do código está coberta por testes unitários	✓	✓
Recuperação de relatórios de execuções de integrações anteriores	x	✓

Estes resultados indicam que o processo de integração contínua implantado diminuiu o tempo gasto com o *deploy* e garantiu atividades que antes não eram executadas como execução automatizada de testes, relatório automático com o resultado das integrações, possibilidade de pessoas alheias aos projetos terem acesso aos

resultados das integrações e emissão de alertas por email. O processo tem o nível de confiabilidade em um percentual aproximado de 90%. Isto infere dizer que o processo está funcional, as tecnologias adotadas atendem as necessidades do processo de produção da empresa, o grau de qualidade dos produtos da empresa se manteve e os envolvidos nos projetos estão satisfeitos com os resultados obtidos.

4. Conclusão

O objetivo desta pesquisa é analisar o processo de produção dos produtos web da empresa Tron Informática Ltda. e implantar a Integração Contínua, visando automatizar este processo afim de diminuir o tempo gasto sem deixar de garantir a alta disponibilidade e a qualidade destes produtos.

De posse dos resultados obtidos e descritos na subseção 3.8, pode-se afirmar que o processo foi implantado com sucesso e que este estudo de caso cumpriu seu objetivo. Todos os passos descritos por Fowler (2006) na subseção 2.4 foram implantados e no momento, os membros da equipe estão integrando seu trabalho diariamente. O processo para realizar *deploy*'s está totalmente automatizado, o que aumentou o desempenho da área de desenvolvimento da empresa.

Sabendo-se ainda que o objetivo da Tron Informática é melhorar todo seu processo de desenvolvimento, alinhando-o aos princípios da Metodologia Ágil, será interessante que estudos futuros sejam realizados para que a empresa tenha um processo de Integração de *Software* Ágil completo. Para isto, orienta-se a aplicação de esforços em trabalhos para implantar as etapas Teste Contínuo (Continuous Testing), Integração de Banco de Dados Contínua (Continuous Database Integration), Inspeção Contínua (Continuous Inspection), Deploy Contínuo (Continuous Deployment), Feedback Contínuo (Continuous Feedback) e Entrega Contínua (Continuous Delivery).

Referências

Lopes, Camilo. (2005) "Integração Contínua com Jenkins", Leanpub.

Rocha, Fábio Gomes. (2016) "Integração contínua: uma introdução ao assunto", http://www.devmedia.com.br/integracao-continua-uma-introducao-ao-assunto/28002, Fevereiro.

Teles, Vinicius Manhães. (2016) "Integração Contínua", http://www.desenvolvimentoagil.com.br/xp/praticas/integracao, Fevereiro.

Pressman, Roger S. (2010), Engenharia de Software. 6. ed. Porto Alegre.

Sommerville, Iam. (2007). Engenharia de software. 8. ed. São Paulo: Pearson Addison.

Gerhardt, Tatiane Engel e Silveira, Denise Tolfo. (2009), Métodos de Pesquisa. Universidade Federal do Rio Grande do Sul. Porto Alegre - RS

Doxey, Jaime Roy. (2011). Metodologia da Pesquisa Científica. Escola Superior Aberta do Brasil. Vila Velha - ES.

Fagundes, Priscila Bastos. (2005). Framework para Comparação e Análise de Métodos Ágeis. Dissertação de Mestrado. Universidade Federal de Santa Catarina. Florianópolis.

Fowler, Martin. (2006). Continuous Integration, http://www.martinfowler.com/articles/continuousIntegration.html, Maio.

Smart, John Ferguson. (2011). Jenkins, The definitive Guide. Creative Commons Edition. O'Reilly Media, 1st edition.

Duvall, Paul M e Matyas, Steve. (2007). Continuous Integration: improving software quality and reducing risk. 1 ed. United States, Pearson Addison.

Jenkins, (2016). Using Jenkins. n.d., https://wiki.jenkins-ci.org, Março.

Jetbrains, (2016). TeamCity Features. n.d., http://www.jetbrains.com/teamcity/features, Março.

CruiseControl, (2016). CruiseControl, a Continuous Integration Toolkit. n.d. http://cruisecontrol.sourceforge.net/index.html, Março.

Chacon, Scoot e Straub, Ben. (2014). Pro Git, everything you need to know about Git. 2 ed. Apress IT eBooks, https://git-scm.com/book/en/v2t, Março.

Goméz, Omar Salvador. (2008). Integración continua en componentes EJB, Escuela Superior Politécnica De Chimborazo, Facultad de Informatica y Electronica.

Viana, Alberto Guimarães e Araujo, José Marcelo Pereira. (2011). Integração Contínua no Processo de Desenvolvimento de Software. Revista Tecnologias em Projeção. vol. 2 - p. 58-59. Junho, 2011.

Riscos Associados à Gestão de Pessoas em Organizações Produtoras de Software

Sérgio Ferreira Maia[1], **Juliano Lopes de Oliveira**[1], **Adriana Silveira de Souza**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Qd D, Câmpus Samambaia
74001-970 – Goiânia – GO – Brasil

fmaia.sergio@gmail.com,{juliano,adriana}@inf.ufg.br

Abstract. *Software organizations face great difficulties in developing and evolving software. Several of these problems are caused by issues related to the management of people involved in software process. Despite the major impact of these issues on software quality, little research has been conducted on the subject, so that the same problems are repeated in different organizations. This paper summarizes and organizes available knowledge on the management of people in software development environments to deal with recurring problems that undermine the productivity of software teams. The proposals presented here were compiled from a review of the literature combined with the observation of development environments in several Brazilian software organizations.*

Resumo. *Organizações produtoras de software (OPS) enfrentam grandes dificuldades para desenvolver e evoluir software. Várias dessas dificuldades são causadas por questões relacionadas à gestão das pessoas envolvidas nos processos de software. Apesar do grande impacto dessas questões nos resultados e na qualidade do software, pouca pesquisa tem sido realizada sobre o tema, de modo que as mesmas dificuldades se repetem em diferentes OPS. Este trabalho sintetiza e organiza o conhecimento disponível sobre a gestão de pessoas em ambientes de desenvolvimento de software de forma a apoiar OPS na abordagem de problemas recorrentes e que prejudicam a produtividade de suas equipes de software. As propostas aqui apresentadas foram compiladas a partir de uma revisão da literatura combinada com a observação de ambientes de desenvolvimento em diversas OPS brasileiras.*

1. Introdução

Grande parte do valor agregado pelos Sistemas de Informação provém do conhecimento sobre dados e processos de negócio codificados em software. No entanto, o custo de construção e manutenção representa uma barreira para a indústria de software. De fato, as atividades envolvidas nesses processos são intelectualmente intensivas, exigindo a manipulação de conceitos abstratos que envolvem habilidades de raciocínio para a resolução de problemas complexos. Dessa forma, as pessoas envolvidas na produção de software precisam despender esforços cognitivos contínuos para dominar uma ampla gama de conhecimentos necessários à construção de software [15].

Um fator que dificulta ainda mais o trabalho relacionado a software é a rápida evolução da tecnologia da informação. Novas técnicas de desenvolvimento surgem a cada

dia e tecnologias se tornam obsoletas em curtos intervalos de tempo. Dessa forma, a vida útil de produtos ligados à tecnologia da informação é muito pequena e as Organizações Produtoras de Software (OPS) precisam lidar estrategicamente com essa busca incessante por atualização tecnológica. Neste sentido, destaca-se a importância de talentos humanos para OPS. O contínuo desenvolvimento de capacidades e habilidades de seus colaboradores é fundamental para a qualidade dos SI desenvolvidos, mas os desafios não se limitam à gestão de conhecimentos na organização. Muitos desafios aparecem mesmo quando as pessoas são competentes no domínio das tecnologias. Logo, é preciso abordar a gestão de pessoas na área de software de uma maneira holística, que trata não apenas do conhecimento de tecnologias, mas abrange as competências cognitivas e comportamentais necessárias ao profissional da área de software [10].

Apesar de todos os desafios envolvidos, existem poucos resultados de pesquisa na gestão de pessoas com foco em equipes de software. Alguns modelos de qualidade, como o P-CMM [2] e o MR-MPS-RH [1], buscam contribuir para o desenvolvimento de uma força de trabalho eficiente com base em estruturas de níveis de maturidade e capacidade para processos organizacionais. Porém, as propostas apresentadas por esses modelos são de difícil implantação, principalmente em organizações de pequeno e médio porte. Mesmo o modelo MPS.BR-RH, que tem como objetivo atender organizações brasileiras desse porte, ainda não oferece orientações ou diretrizes para orientar o trabalho de implementação de suas propostas. Dessa forma, implementar as práticas de gestão de pessoas recomendadas pelos principais modelos de referência disponíveis na atualidade constitui um grande desafio, principalmente para OPS iniciantes, que buscam atingir os primeiros níveis de maturidade na gestão de pessoas.

O presente artigo contribui para uma mudança nesse cenário, apresentando um conjunto específico de propostas para abordar problemas recorrentes que dificultam a formação e a aplicação de talentos humanos nas OPS brasileiras. O trabalho oferece às OPS orientações direcionadas para a implementação de boas práticas de gestão de pessoas, apoiando a definição de estratégias adequadas para o estabelecimento de ambientes apropriados para a realização de processos de software. Essas propostas foram sintetizadas a partir de uma metodologia que combinou uma pesquisa descritiva (conduzida na forma de uma revisão da literatura, com o objetivo de analisar os principais problemas relacionados ao desenvolvimento de software e compreender a sua correlação com a gestão de pessoas) com uma análise de cenários de diversas OPS visitadas no contexto de ações de melhoria de processos de software (implementação e avaliação de modelos de qualidade de processo e de produto conduzidas pelos pesquisadores). Nessas ações as práticas de gestão de pessoas nas OPS foram observadas para identificar problemas e abordagens relacionadas a essa gestão. Com base na combinação dos relatos da literatura com as práticas observadas na indústria de software brasileira, foram compilados os principais problemas e soluções relacionados à gestão de pessoas em OPS brasileiras.

O restante deste artigo discute e caracteriza os problemas identificados na forma de riscos associados à gestão de pessoas em equipes de software. Para cada risco são analisadas suas possíveis causas, as consequências da sua ocorrência e possíveis abordagens para sua mitigação ou contingência em OPS. A Seção 2 discute a fundamentação teórica do trabalho, que envolve os modelos existentes para gestão de pessoas em OPS. A Seção 3 apresenta a síntese de riscos relacionados com a gestão de pessoas obtida da

pesquisa realizada. A Seção 4 discute os resultados e as contribuições dessa pesquisa e apresenta direções para trabalhos futuros.

2. Gestão de Equipes de Software

A gestão de pessoas é o processo que as organizações utilizam para gerenciar o comportamento de seus colaboradores. O processo orienta os gestores sobre como gerir as pessoas que a eles respondem. Como o comportamento humano influencia diretamente a qualidade de produtos e serviços de uma organização, a gestão de pessoas tem sido aplicada com sucesso para fomentar conhecimentos, competências e maturidade dos recursos humanos em diversos tipos de organizações [14]. No entanto as OPS possuem idiossincrasias que exigem abordagens mais específicas para o processo de gestão de pessoas. A cultura organizacional, por exemplo, é um fator crítico para o sucesso de iniciativas de melhoria de processos em OPS [6]. A natureza puramente abstrata do software também dificulta a gestão do trabalho em OPS, tornando o planejamento e o acompanhamento de tarefas relacionadas à produção de software consideravelmente mais complexos.

Apesar dessas dificuldades, apenas dois modelos identificados na literatura apresentam propostas para abordar as especificidades da gestão de pessoas em equipes de produção de software. Esses modelos de qualidade, que visam aprimorar a capacidade de gestão de pessoas envolvidas com software, ainda são pouco conhecidos pelas OPS brasileiras. O Modelo MR-MPS-RH [1], embora descreva níveis de maturidade e resultados esperados que indicam uma possível sequência para a implementação da gestão de pessoas, ainda é incipiente. O modelo disponibiliza apenas um guia geral de processos e resultados, sem discussões sobre como aplicar esses processos e resultados em uma OPS. Logo, o modelo não auxilia as organizações no planejamento de ações efetivas para alcançar os níveis de maturidade propostos. Já o Modelo P-CMM [2] apresenta uma discussão profunda e abrangente sobre práticas recomendadas para a gestão de pessoas em OPS. O modelo define cinco níveis de maturidade para essa gestão e distribui as práticas recomendadas em vinte e duas áreas de processo, como mostra a Figura 1.

No entanto, as OPS têm dificuldade para usar o modelo P-CMM devido à complexidade dos processos e práticas propostos, bem como ao volume de informações apresentadas sobre cada aspecto tratado no modelo. OPS que iniciam sua busca pela maturidade do processo de gestão de pessoas dificilmente conseguem assimilar e implementar o vasto conjunto de propostas descritas no modelo P-CMM, mesmo considerando apenas o primeiro nível de maturidade proposto no modelo (que é o nível II, Gerenciado, já que o nível I do modelo indica total falta de maturidade no processo de gestão de pessoas).

Além disso, a estratégia de implementação proposta pelo P-CMM é planejada para que os elementos do modelo funcionem conjuntamente. Cada componente do P-CMM está entrelaçado a vários outros, formando uma complexa rede de dependências. Uma evidência disso são os *threads* de áreas de processo previstos no modelo [2]. Isso acontece com componentes de todos os níveis do modelo. Estruturas maiores, como os próprios níveis de maturidade, também apresentam características similares: cada nível está fortemente ligado ao nível anterior, e sofre fortes influências das áreas de processo anteriores. Logo, é muito difícil tratar uma ou mais áreas de processo do modelo de forma isolada, de forma que o modelo precisa ser implementado em sua totalidade, o que não é viável para OPS de pequeno porte.

Figura 1. Áreas de Processo da Gestão de Pessoas segundo o P-CMM

Assim, a motivação para a pesquisa aqui descrita é a falta de apoio para OPS iniciantes e de porte reduzido na adoção de boas práticas de gestão de pessoas. A ideia é disponibilizar um conjunto de ações simples e diretas, que facilite a implementação dessas práticas, considerando as limitações de recursos existentes na maioria das OPS brasileiras.

3. Riscos para a Gestão de Pessoas

Esta seção apresenta uma proposta para disseminar boas práticas de gestão de pessoas, viabilizando sua aplicação em qualquer tipo de OPS. A pesquisa realizada identificou dificuldades recorrentes enfrentadas pelas equipes técnicas em OPS brasileiras. Essas dificuldades são aqui conceituadas e caracterizadas na forma de riscos, com a identificação de possíveis causas, a discussão de consequências da sua ocorrência e a indicação de abordagens para mitigação ou contingência de cada risco na organização produtora de software. A Tabela 1 apresenta uma síntese dos principais riscos identificados.

Cada organização pode determinar quais riscos são mais impactantes ou mais prováveis de ocorrer dentro de sua realidade de trabalho. Dessa forma, a organização pode montar uma estratégia adequada à sua realidade, em vez de seguir um conjunto de práticas predefinidas por um modelo de qualidade genérico, cuja implementação pode ser inviável, principalmente para OPS de pequeno porte.

3.1. Estresse

O estresse é uma característica evolucionária que contribui para a sobrevivência de espécies em adaptação a ambientes hostis. Os reflexos do estresse são resultantes de reações físicas e psicológicas de um indivíduo submetido a situações que exigem grande esforço para adaptação. O *distresse* é uma forma negativa de estresse, marcada por sintomas físicos e psicológicos desagradáveis resultantes de uma situação difícil, que exige

Tabela 1. Síntese de Riscos para a Gestão de Pessoas

Risco e **Possíveis Causas**	**Possíveis Consequências**	**Possíveis Abordagens**
Estresse: Ambiente ruim; pressão; sobrecarga de trabalho; insatisfação salarial; organização instável; sensação de insuficiência; falta de reconhecimento	Absenteísmo; presenteísmo; improdutividade; alta rotatividade; gastos com problemas de saúde dos colaboradores	A solução para este risco envolve conhecer e mitigar todos os demais riscos
Rotatividade de pessoal: Fenômenos internos (política salarial, falta de plano de carreira, ambiente inadequado); e externos (oferta e procura de capital humano, macro-economia)	Dificuldade para reter talentos; força de trabalho sem competência para atuar no seu mercado	Motivação; desenvolvimento profissional; ambiente de trabalho saudável; cultura organizacional participativa
Preconceitos: Diferenças e diversidade da sociedade; intolerância e ignorância	Baixa produtividade; estresse; desconfiança na organização; insegurança na própria competência	Gestão estratégica de pessoas com foco na diversidade
Falhas na comunicação: Ruído; falta de treinamento; diferenças intelectuais; falha de envio e recepção de dados; ambiente de transmissão inadequado; conflitos de linguagem	Projetos inacabados e abandonados; atritos; gestão ineficaz; retrabalho; atrasos; custos extras; clientes e funcionários insatisfeitos; descrédito	Desenvolver habilidades de comunicação dos colaboradores; construir meios eficazes de transmissão de informação
Falhas de liderança: Despreparo para assumir o cargo; falta de treinamento e capacidade adequada; estagnação temporal dos profissionais	Conflitos; inibição da inovação e criatividade; estresse; problemas de autoestima; desmotivação; rotatividade	Definição de responsabilidades dos gestores; melhoria de maturidade e capacidade de líderes; gestão de pessoas
Ambiente tóxico: Ambiente físico ruim; falta de oportunidades; políticas injustas; intrigas; falhas de comunicação; excesso de trabalho; falha de liderança	Sobrecarga de trabalho de alguns; ociosidade de outros; desânimo; conflitos; alta rotatividade de pessoal	Planejar o ambiente de trabalho, sem distrações e com recursos necessários para realizar as atividades
Gestão ineficaz: Demanda reprimida; custo da manutenção; dificuldades com prazo; planejamento ruim; processos ineficientes; incompetência gerencial	Insatisfação das equipes; prejuízos por retrabalho; insatisfação de clientes; prazos e custos altos; problemas judiciais	Recursos e força de trabalho disponíveis; uso de boas práticas de gestão; capacitação de gestores de projeto
Processos ineficientes: Equipe imatura; responsabilidades indefinidas; desalinhamento com o negócio; objetivos instáveis; burocracia; descontrole	Manutenção mais cara; conflitos nas equipes; problemas com custos e prazos; insatisfação de clientes e funcionários	Treinamento contínuo; *swap* controlado; evolução contínua de tecnologias e processos

mais do que o indivíduo pode ou quer suportar. No âmbito organizacional, a variante mais comum é o estresse ocupacional e indica que o ambiente de trabalho está forçando o colaborador a passar por situações desagradáveis, e que este não está preparado psicologicamente para superá-las e internalizá-las positivamente [3].

Os fatores que levam ao estresse ocupacional podem ser classificados em físicos, sociais e emocionais. As causas geradoras de fatores estressores são inúmeras e estão ligadas a fatores cognitivos, pois cada pessoa reage de maneiras diferentes a uma mesma situação: descontentamento com colegas, sobrecarga de trabalho, ambiguidade de prioridades, pouco tempo e recursos para a realização das atividades, insatisfação salarial, instabilidade no emprego, sensação de insuficiência profissional, pressão por eficiência, falta de reconhecimento, entre outros fatores [4].

O estresse causa consequências negativas à organização, pois diminui a produtividade, aumenta a rotatividade e também os gastos oriundos de problemas de saúde da sua força de trabalho. De fato, há indícios de que o estresse está presente na maioria das organizações brasileiras, como demonstra a pesquisa descrita em [22]. A pesquisa envolveu 547 gestores de 15 empresas brasileiras de diferentes ramos da economia e constatou que 63% dos gestores apresentavam estresse. Destes, 45% apresentavam nível de estresse leve ou moderado, 15% apresentavam nível de estresse intenso e 3% apresentaram um nível muito intenso.

Funcionários estressados são, em geral, consequência de outros problemas presentes na organização. Assim, a solução para o estresse passa pela extinção dos vetores que causam esse problema na organização. Dentre esses vetores se destacam: rotatividade de pessoal, ambiente de trabalho tóxico, preconceitos, falha de comunicação, gestão incompetente, controle de projetos ineficiente e processo de desenvolvimento problemático. As próximas seções discutem esses problemas e propõem formas para minimizá-los ou eliminá-los, de forma a diminuir a incidência de estresse na força de trabalho.

3.2. Rotatividade de Pessoal

A rotatividade de pessoal (ou *Turnover*) representa o movimento de entrada e saída de colaboradores de uma organização, sendo uma das principais responsáveis pela dinâmica organizacional. Quando a rotatividade de pessoal acontece de forma consciente e controlada, ela se torna uma ferramenta necessária e útil para a organização. No entanto, quando não controlada, a rotatividade representa um obstáculo para qualquer tipo de organização, implicando em muitos custos e desvantagens competitivas. O descontrole da rotatividade em uma organização é fruto de uma gama de outros problemas, ou seja, a rotatividade não é uma causa, mas uma consequência de certos fenômenos localizados interna ou externamente à organização [5].

Um cenário de alta rotatividade pode gerar vários efeitos colaterais indesejáveis para a organização, incluindo perda de produtividade e lucro, degradação do clima organizacional, desmotivação e falta de compromisso por parte das pessoas e até mesmo interferindo na credibilidade e imagem da organização perante o mercado. Um índice alto e descontrolado de rotatividade indica que a organização não consegue reter pessoas de talento. Para combater tal situação é necessário um conjunto de ações para motivar o colaborador. Isso compreende definir maneiras para o desenvolvimento pessoal e profissional dos funcionários, criar um ambiente de trabalho saudável e manter um bom

relacionamento entre todos os colaboradores da organização. Ações mais específicas estabelecem punição e compensação justas para os colaboradores e uma cultura participativa na organização. Assim, a organização precisa perceber a necessidade de investir em seus funcionários como um ativo organizacional para alcançar seus objetivos de negócio. O amadurecimento da cultura organizacional cria uma imagem de respeito e interesse por parte dos profissionais e diminui a rotatividade de pessoal.

3.3. Preconceitos

Preconceito é um juízo preconcebido que se manifesta como uma atitude discriminatória perante pessoas. É uma ideia formada antecipadamente, sem uma base de informações ou conhecimentos que possibilitem uma conclusão fundamentada em argumentos verídicos ou raciocínio lógico. O preconceito é fruto de estigmas que surgiram na sociedade, visando vantagens políticas e econômicas. Embora não sejam mais aplicados à nova realidade social, esses tabus estão enraizados na cultura global e se manifestam desde simples pensamentos até agressões físicas e verbais. A diversidade pode ser conceituada como um conjunto de características, visíveis ou não, que diferem as pessoas umas das outras [9]. Dentro da diversidade ocorrem grupos, chamados de minorias, que são os principais alvos do preconceito. [18] propõe modalidades ou formas de preconceitos advindos das diversidades humanas:

- O preconceito intelectual é identificado em situações onde pessoas discriminam outrem por questões relacionadas às suas capacidades cognitivas, quantidade de títulos acadêmicos ou nível de conhecimento teórico.
- O preconceito etário é oriundo de relações sociais entre grupos de diferentes faixas de idade. Uma de suas formas mais comuns é o preconceito em relação aos idosos.
- O sexismo é uma atitude, discurso ou comportamento que se baseia na discriminação sexual. Geralmente ocorre de pessoas do sexo masculino para com pessoas do sexo feminino, mas também pode ocorrer o inverso.
- A homofobia denota comportamentos discriminatórios e até violentos contra pessoas com orientação sexual homoafetiva, ou que apresentem identidade sexual disforme do aceito pela sociedade.
- O preconceito racial se identifica na discriminação ou segregação relacionadas às características de raça de um determinado grupo, ao passo que o preconceito étnico demonstra conflitos vindos de choques culturais entre diferentes povos.
- O preconceito de credo está ligado ao lado emocional do ser humano e envolve a divergência sobre questões de religião, metafísicas e ideológicas.
- O preconceito de nacionalidade surge de diferenças culturais ou físicas, relacionadas ao patriotismo, costumes ou mesmo o idioma de comunicação de um país.
- O preconceito de deficiência surge contra pessoas com necessidades especiais. Essas pessoas encontram grande dificuldade em se inserir no mercado de trabalho, pois o preconceito contra elas aparece em um nível mais amplo, revelado na falta de acessibilidade, no sistema educacional mal preparado para atender suas necessidades especiais e nas atitudes manifestas pelas pessoas ditas normais [17].

Existem vários benefícios ligados à adoção de uma postura organizacional acolhedora quanto às minorias, entre eles: atração de talentos, conquista de diferentes segmentos de mercado, flexibilidade organizacional e valorização da imagem perante a sociedade, que se sensibiliza pela causa das minorias [19]. Problemas relacionados a preconceitos,

discriminação e intolerância causam gastos à organização, afetam o bem-estar e a produtividade dos funcionários, geram estresse, aumentam os estereótipos e causam perda de confiança na própria competência dos colaboradores [13]. Para combater o preconceito nas organizações é necessária uma gestão das diversidades na organização, visando evitar situações negativas relacionadas ao tema. Essa gestão pode ser alcançada com um enfoque adequado em atividades de alguns setores da organização, como o de recrutamento e seleção, treinamento e comunicação [8].

3.4. Falhas na comunicação

O processo de comunicação envolve um comunicador e um conjunto de receptores em uma relação dinâmica e interativa, na qual um influencia o outro através de um meio (fala, sinais, livro, TV, internet e outros) onde as mensagens são enviadas e recebidas [13]. Na comunicação dentro das organizações o conteúdo das mensagens compartilhadas faz parte do cotidiano do trabalho realizado. Uma organização possui vários processos onde grupos de pessoas trabalham em conjunto com a finalidade de alcançarem um objetivo em comum. Nesse ambiente, a transmissão eficiente de informações é importantíssima, já que um déficit na comunicação pode causar enormes prejuízos. Por isso, a falha de comunicação é um problema cada vez mais impactante nas organizações.

Falha na comunicação é um fator gerador de problemas em projetos, estando entre as principais razões pelas quais um projeto não atinge seus objetivos [16]. Existem vários fatores que levam à comunicação ineficiente e não confiável dentro de uma organização, tais como: o meio utilizado para a transmissão de mensagens, as diferenças intelectuais, profissionais e de valores, o ruído (causado pelos meios de comunicação ou pelos próprios comunicadores) e a falta de treinamento. A falha de comunicação causa gastos extras de recursos e tempo, aumento na complexidade do trabalho, conflitos entre equipes, descontentamento e não atendimento das expectativas dos clientes diante da falta de qualidade de produtos e serviços.

Para solucionar ou minimizar os problemas causados por uma comunicação ineficiente, é necessário uma gestão intensa dos processos de distribuição de informações dentro de uma organização. Na gestão de projetos organizacionais, em particular, a importância da qualidade da comunicação é comprovada [16]. Para evitar problemas de comunicação a organização precisa garantir a geração, coleta, distribuição, armazenamento, recuperação e destinação final das informações sobre o trabalho realizado na organização de forma oportuna e adequada.

3.5. Falhas de liderança

A liderança está ligada à habilidade de direcionar, desenvolver e instigar um grupo de pessoas, tornando-as mais eficientes, criativas e produtivas e ajudando-as a alcançar um objetivo comum. O conceito de liderança está fortemente relacionado à estrutura organizacional e influencia fortemente o sucesso ou fracasso dessa estrutura. A liderança não é um dom inerente ao ser humano, pois um líder não terá sucesso quando inserido em qualquer grupo de pessoas ou em qualquer contexto de trabalho. Diferentes grupos exigem abordagens de liderança distintas. Uma postura errada de um líder perante sua equipe traz vários prejuízos ao grupo. Além disso, o despreparo para assumir o cargo, a falta de treinamento e de capacidade adequados e a estagnação temporal dos profissionais acabam

criando *líderes tóxicos* nas organizações. Pode-se citar como condutas problemáticas a liderança superprotetora, o microgerenciamento e o subgerenciamento [7].

As dificuldades que os profissionais têm em exercer suas atividades de liderança são consequências de vários fatores, como perfil inapropriado para a função, falta de treinamento adequado, tempo limitado para realizar o trabalho e medo de punição ao executar certas tarefas [20]. A organização que deseja melhorar sua liderança precisa realizar uma ação conjunta entre líder e liderados para ajudar o líder a desenvolver as habilidades e competências necessárias para guiar o desenvolvimento da equipe. Deve-se mostrar aos líderes (chefes ou gerentes) as suas responsabilidades com a sua força de trabalho e ajudá-los na evolução de sua capacidade de gerir e de sua maturidade para que eles tomem para si suas responsabilidades [2].

3.6. Ambiente de trabalho tóxico

O ambiente de trabalho compreende tudo aquilo que rodeia as pessoas enquanto estas, na qualidade de colaboradoras, executam suas atividades produtivas para uma organização. Esse ambiente envolve não só condições físicas de higiene e segurança, mas também relações sociais, culturais e econômicas que influenciam o trabalho da organização. Um ambiente de trabalho tóxico contém vários dos problemas discutidos nas seções anteriores e resultam em um local hostil, onde o funcionário enfrenta dificuldades para realizar suas tarefas e não vislumbra oportunidades para crescimento profissional e pessoal.

A falta de qualidade do ambiente de trabalho dentro de uma organização está ligada a vários fatores físicos e psicossociais, gerando uma série de consequências negativas e deteriorando a saúde organizacional. Os fatores físicos incluem condições de instalações e equipamentos do local de trabalho, que acabam atrapalhando os colaboradores na realização de suas atividades. Por exemplo, equipamentos escassos ou precários e o espaço físico reduzido ou inadequado para o trabalho leva a situações de estresse, sobrecarga de trabalho ou ociosidade, desânimo e gera conflitos entre os colaboradores. Em OPS, o ruído do ambiente é um dos fatores que mais dificultam a concentração e leva a uma redução da capacidade produtiva e do nível de satisfação dos colaboradores com o seu trabalho. Já os fatores psicossociais estão ligados às dificuldades de relacionamento entre os integrantes do ambiente de trabalho. Muitas dessas dificuldades são oriundas de preconceitos, como já foi discutido, e podem levar à ocorrência de conflitos, desavenças e até mesmo discussões e brigas entre pessoas que compõem a força de trabalho da organização.

Normalmente a qualidade de vida dos colaboradores recebe pouca atenção em OPS de menor porte. No entanto, as consequências de um ambiente de trabalho tóxico tendem a gerar gastos maiores do que aqueles oriundos de uma gestão adequada do ambiente de trabalho, preocupada com o bem-estar e a motivação de seus colaboradores [3]. Para propiciar uma boa qualidade de vida para os colaboradores, o ambiente de trabalho precisa oferecer mais do que segurança e saúde. É necessário associar esse ambiente a uma melhoria contínua do clima organizacional, oferecendo condições psicológicas adequadas aos colaboradores para se desenvolverem. Para profissionais de software, a perspectiva de crescimento na carreira e de aquisição de conhecimentos supera muitas outras preocupações, inclusive as relacionadas à remuneração. Melhorando o ambiente físico e psicossocial, aumenta-se a produtividade, o bem-estar e a segurança, e isso motiva os colaboradores da organização [3].

3.7. Falhas na Gestão de Projetos

Problemas relacionados ao gerenciamento dos projetos atrapalham ou impedem as organizações de alcançarem seus objetivos [21]. Um desafio para a gestão de projetos em OPS é tratar o problema da Demanda Reprimida, que é uma uma fila de espera de sistemas a serem desenvolvidos em uma organização. Ela acontece quando a demanda por sistemas é maior que os recursos disponíveis para construi-los, de modo que a necessidade de novos sistemas cresce mais rápido do que a capacidade da organização de os produzir. Esta situação causa grandes transtornos no planejamento de custos da organização, pois ela tende a avaliar apenas a demanda visível (a parte da demanda futura sobre a qual a organização possui conhecimento). Essa diferença contínua entre o orçamento planejado e os reais gastos faz com que projetos sejam abandonados por falta de recursos.

Outro desafio é tratar o grande volume de manutenção. As organizações gastam muito tempo e recursos na manutenção de seus sistemas, o que contribui para aumentar o problema de demanda reprimida. A atividade de manutenção acaba competindo por recursos com os novos projetos, levando a organização a abandonar a construção de novos sistemas para garantir suporte aos sistemas antigos. Os altos custos do processo de manutenção se explicam pelo próprio sucesso desse processo, que garante longevidade para um grande número de sistemas em funcionamento, que precisam de suporte e evolução [11]. Por outro lado, processos de desenvolvimento inadequados geram software com documentação ruim, o que aumento o esforço de manutenção.

Prazos mal estimados também dificultam a gestão de projetos de software. Sempre que um projeto de software é iniciado, deveria haver planejamento e alocação de recursos e tempo para sua execução. Entretanto, inicialmente apenas parte das informações necessárias para uma estimativa correta está disponível. Situações extraordinárias, mudanças nos requisitos do projeto por parte do cliente e evolução da tecnologia são fatores que influenciam o prazo de construção e que uma equipe não pode prever durante o planejamento. Assim as estimativas de prazo, na maioria das vezes, não são corretas, causando insatisfação dos clientes, pressão sobre a equipe técnica, perda de negócios importantes, abandono ou aumento substancial do custo dos projetos e quebra de contratos.

Os problemas citados possuem uma forte interdependência, de modo que ao se minimizar o impacto de um problema, reduz-se o impacto de muitos outros. Por exemplo, o efeito *iceberg* é uma extensão do problema de demanda reprimida, refletindo o cenário onde o número de demandas invisíveis é maior do que o de demandas visíveis. A demanda reprimida afeta diretamente o tempo necessário para a implementação de um SI, acarretando em entregas fora do prazo. Assim, retirar recursos de atividades de desenvolvimento e direcioná-los para atividades de manutenção não é uma solução adequada, pois agrava o problema da demanda reprimida. Por outro lado, diversas atividades podem ser empregadas para minimizar o impacto de demanda reprimida, efeito *iceberg* e problemas de prazo para a organização, tais como:

- Aumento da força de trabalho por meio de novas contratações: mais funcionários significam mais recursos humanos para produzir, desenvolvendo os projetos assumidos com maior velocidade.
- Investimento em recursos de melhor qualidade: em um ambiente de desenvolvimento de software, usar tecnologia de qualidade pode melhorar a produtividade nos projetos. Um ambiente de trabalho que atente para a segurança e saúde dos

colaboradores, oferecendo-lhes um lugar sem distrações, também contribui para uma maior produtividade.

- Formação de uma força de trabalho mais talentosa: apenas contratar funcionários não é uma solução efetiva para problemas relacionados à produtividade dos funcionários em organizações [5]. Quando a organização já conta com número suficiente de colaboradores para realizar suas tarefas, ela deve investir na maturidade desses profissionais, desenvolvendo suas capacidades, habilidades e competências necessárias para desenvolver softwares de qualidade [10]. Isso contribui para diminuir os prazos de entrega e facilita as atividades de manutenção.
- Aplicação de técnicas de melhoria e qualidade (organizacional e de software): gestão de projetos envolve planejamento, execução e monitoramento do trabalho. Todas essas atividades devem ser realizadas eficientemente para se obter um produto de qualidade e dentro do prazo e custo planejados. A adoção de boas técnicas de desenvolvimento auxiliam a alcançar esses resultados pela diminuição do retrabalho e melhoria da qualidade de software, o que facilita a manutenção.

3.8. Processos ineficientes

O processo de desenvolvimento de software contempla várias atividades, parcialmente ordenadas, que são realizadas por uma equipe de profissionais qualificados, com tempo e recursos predefinidos, visando a construção e manutenção de um sistema [15]. Um processo de qualidade, com um planejamento adequado, definição de um ciclo de vida apropriado e uma gerência eficiente, aumenta a probabilidade de se gerar um produto de qualidade entregue dentro do prazo, e atendendo às expectativas do cliente. Entretanto, várias são as dificuldades encontradas no processo de desenvolvimento de sistemas na maioria das OPS. São problemas advindos da ineficiência na definição do planejamento de projeto, mas também acontecem por outros motivos, tais como:

- Diferença de conhecimento: o processo de um software é organizado em etapas parcialmente ordenadas. Se uma etapa inicial não é eficientemente realizada pela equipe responsável, todas as outras etapas sofrerão as consequências disso, não importando o quão bons sejam os funcionários responsáveis por elas. Para evitar isso, é necessário um alinhamento do desempenho da força de trabalho na organização, homogeneizando a capacidade de indivíduos, grupos e unidades organizacionais. Assim, as etapas de trabalho devem ser realizadas por pessoas igualmente capacitadas.
- Relacionamento com o cliente: geralmente clientes não possuem conhecimento técnico na área de software para compreender os limites e dificuldades da construção de seus sistemas, o que leva a problemas de comunicação com a equipe responsável pelo projeto. Realizar constantes reuniões com o cliente, inserindo-o no projeto e mantendo contato durante todo o processo de desenvolvimento ajuda a evitar esses problemas. Além disso, é necessário treinamento para desenvolver as habilidades dos profissionais responsáveis pela mediação entre clientes e equipes técnicas.
- Engenharia de requisitos: a engenharia de requisitos é um processo que visa à identificação das necessidades do cliente a fim de encontrar uma definição completa e correta do sistema que o atende. Esse processo é altamente sujeito a erros, tais como requisitos errados ou incompletos, mudança de requisitos por

parte do cliente e requisitos conflitantes. Por ser uma das etapas mais importantes do processo de desenvolvimento de um sistema, é necessário que as atividades inerentes a ela sejam realizadas por uma equipe capacitada. A habilidade de comunicação aliada a um treinamento com foco no conhecimento necessário para a boa execução dessas tarefas pode ajudar.

- Troca de contexto: *swap* ou troca de contexto trata de mudança de função de um profissional entre um projeto e outro, ou dentro de um mesmo projeto. Se equipes são modificadas constantemente, não haverá coesão. Se o colaborador troca de contexto continuamente, ele perde seu foco. Consequentemente, os grupos nunca alcançam um trabalho de equipe eficiente. Além disso, a organização deixa de usufruir da total capacidade de funcionários competentes em determinada área específica. O *swap* é algo interessante para a organização quando feito de forma estratégica. Se torna interessante trabalhar com movimentação de funcionários entre equipes dentro da organização, seguindo planos estratégicos definidos para capacitar e desenvolver os talentos humanos da organização. Com a criação e eliminação contínua de grupos, a sinergia aumenta na organização e as habilidades de comunicação e de coordenação de tarefas são evoluídas.
- Código sujo e falta de documentação: o desenvolvimento de software deve visar à manutenção do sistema. Seguir boas práticas de programação a fim de manter um código limpo, de fácil leitura e eficiente diminuem os gastos com manutenções e correções de erro futuras [12]. Além disso, a documentação é de extrema importância para o software, cuja natureza abstrata torna-o difícil de visualizar e gerenciar. Sem um conjunto de informações documentadas sobre o software, não é possível avaliar o progresso do projeto em um momento qualquer do processo de desenvolvimento do software. As tarefas de desenvolver código limpo e gerar boa documentação, quando recebem a devida atenção, geram software de melhor qualidade e diminuem custos de manutenção. Essas duas tarefas devem ser tratadas como objetivos de negócio da organização. Com isso, deve haver investimento em treinamento de pessoal e na criação de processos bem definidos para realizar uma documentação adequada e padronizada nos projetos de software.
- Técnicas ultrapassadas: apesar da evolução constante das técnicas de engenharia de software, muitas organizações continuam realizando seus processos de maneira amadora. A natureza humana teme o novo, e a cultura das organizações não é diferente, já que elas são formadas por pessoas. Técnicas e normas de desenvolvimento ultrapassadas ou desnecessárias atrapalham a produtividade das OPS, dificultam o processo de desenvolvimento, e confundem os profissionais. Além disso, são altamente desmotivadoras, aumentando o estresse no trabalho e a rotatividade espontânea. Para evitar esse cenário a organização deve investir esforços para aplicar metodologias e tecnologias de desenvolvimento de software atuais, mas adequadas à sua realidade. Ela também deve buscar implementar um programa de melhoria de processos, para assegurar a evolução da sua forma de trabalho visando atender seus clientes da melhor maneira possível.

4. Conclusões

A principal matéria-prima da indústria de software é o conhecimento humano. A pesquisa descrita neste artigo mostra que a origem de muitos problemas dessa indústria está na estrutura deficiente de gestão de pessoas e na baixa maturidade de sua força de trabalho.

Apesar de existirem modelos de qualidade que visam aprimorar a capacidade de pessoas envolvidas com processos de software, esses modelos ainda são pouco conhecidos e pouco aplicados nas OPS brasileiras, principalmente porque não apresentam abordagens viáveis para OPS de pequeno porte.

Este artigo identifica e discute dificuldades recorrentes enfrentadas para a gestão de equipes técnicas em ambientes de desenvolvimento de software. Essas dificuldades são conceituadas e caracterizadas na forma de riscos, com a identificação de possíveis causas, a discussão de consequências da sua ocorrência e a indicação de abordagens para mitigação ou contingência de cada risco na organização produtora de software.

Um diferencial da proposta baseada em riscos apresentada neste artigo com relação aos modelos de qualidade de gestão de pessoas baseados em níveis de maturidade, como o P-CMM e o MR-MPS-RH, é que cada organização pode avaliar o conjunto de riscos proposto e decidir quais são os mais impactantes ou mais prováveis dentro de sua realidade. Assim, a organização tem subsídios para definir uma estratégia de gestão de pessoas adequada às suas necessidades. Não há um conjunto predefinido de práticas e processos obrigatórios, e isso torna a abordagem viável, mesmo para OPS de pequeno porte e iniciantes nos processos de gestão de pessoas.

As propostas para contingência e mitigação de riscos apresentadas neste artigo foram comparadas com as do modelo P-CMM, e constatou-se que algumas áreas de processo desse modelo não são aplicadas para minimizar qualquer dos problemas identificados neste artigo. Isso pode ser um indício de que ainda há problemas a serem identificados em ambientes de desenvolvimento de OPS brasileiras, ou que essas áreas de processos abordam situações pouco comuns nessas organizações. Essa investigação está sendo conduzida como continuação do trabalho de pesquisa descrito neste artigo.

Outra ação em andamento para a continuidade desse trabalho é a realização de uma pesquisa de campo para avaliar junto às OPS a relevância dos riscos identificados e das propostas para sua abordagem. Nesse sentido, um questionário já foi elaborado e deve ser distribuído para a coleta de informações. Além disso, como trabalho futuro, experimentos precisam ser conduzidos para obter evidências de que as abordagens descritas para contingenciar e mitigar os riscos, obtidas de forma isolada na literatura e na observação de boas práticas de OPS, podem funcionar sinergicamente quando aplicadas de forma conjunta.

Referências

[1] Associação para Promoção da Excelência do Software Brasileiro. *Guia Geral MPS de Gestão de Pessoas*. SOFTEX, Campinas, SP, Agosto 2016.

[2] B. Curtis and B.Hefley and S. Miller. *People capability maturity model (P-CMM) - version 2.0 - Technical Report CMU/SEI-2009-TR-003*. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2009.

[3] M. R. P. Benke and E. Carvalho. Estresse x qualidade de vida nas organizações: um estudo teórico. *Revista Objetiva-Rio Verde*, 2008. Disponível em: http://www.faculdadeobjetivo.com.br/arquivos/Estresse.pdf.

[4] S. H. H. Camelo and E. L. S. Angerami. Symptoms of stress in workers from five family health centers. *Revista latino-americana de enfermagem*, 12(1):14–21, 2004.

[5] I. Chiavenato. *Gestão de pessoas: O novo papel dos recursos humanos nas organizações*. Rio de Janeiro: Elsevier, 2004.

[6] P. G. Fernandes and J. L. de Oliveira. Perfil cultural e institucionalização de processos de software: Estudo de caso em duas organizações de software. In *VI Workshop Olhar Sociotécnico sobre a Engenharia de Software (anais do Simpósio Brasileiro de Qualidade de Software)*. SBC, Junho 2010.

[7] J. O. Fiorelli. *Psicologia para administradores: integrando teoria e prática*. Atlas, 9ª edition, 2014.

[8] M. T. L. Fleury. Gerenciando a diversidade cultural: experiências de empresas brasileiras. *Revista de Administração de Empresas*, 40(3):18–25, 2000.

[9] A. S. Godoy, M. Texeira, and L. ZACCARELLI. *Gestão do fator humano: uma visão baseada nos stakeholders*. Saraiva, São Paulo, SP, 2008.

[10] IEEE Computer Society. *Software Engineering Competency Model version 1.0 (SWECOM)*. IEEE, 2014.

[11] M. M. Lehman and J. F. Ramil. Rules and tools for software evolution planning and management. *Annals of Software Engineering*, 11(1):15–44, Nov 2001.

[12] R. C. Martin. *Agile software development: principles, patterns, and practices*. Prentice Hall PTR, 2ª edition, 2003.

[13] G. d. C. Martins and A. Garcia. Conflito interpessoal entre brasileiros e entre brasileiros e estrangeiros em empresas multinacionais de manaus, am. *Cadernos de Psicologia Social do Trabalho*, 14(2):179–194, 2011.

[14] M. S. Pacheco. Evolução da Gestão de Recursos Humanos: um estudo de 21 empresas. Master's thesis, USP - Faculdade de Economia, Administração e Contabilidade, Ribeirão Preto, SP, 2009.

[15] Pierre Bourque and Richard Fairley, editor. *SWEBOK v3.0 - Guide to the Software Engineering Body of Knowledge*. IEEE, 2014.

[16] Project Management Institute. *A Guide to the Project Management Body of Knowledge: PMBOK Guide*. PMI, 5 edition, 2013.

[17] R. P. D. Ribeiro and M. E. A. Lima. O trabalho do deficiente como fator de desenvolvimento. *Cadernos de Psicologia Social do Trabalho*, 13(2):195–207, 2010.

[18] S. P. Robbins. *Fundamentos do Comportamento Organizacional*. Saraiva, São Paulo, SP, 7ª edition, 2003. Traduzido por Reinaldo Marcondes.

[19] R. G. d. Santos and L. Diehl. Diversidade humana em organizações de lajeado/rs: Desafios e tendências para a gestão de pessoas. *Estudo & Debate*, 20(2), 2013.

[20] B. Tulgan. *Não tenha medo de gerenciar seu chefe*. Sextante, 1ª edition, 2012.

[21] R. V. Vargas. *Gerenciamento de Projetos*. Brasport, 6ª edition, 2005.

[22] L. Zille. Novas perspectivas para a abordagem do estresse ocupacional em gerentes: estudo em organizações brasileiras de setores diversos. *Belo Horizonte: CEPEAD/UFMG*, 2005.

Implantação de um Service Desk com Apoio da Ferramenta de Software Livre GLPI em um Ambiente Hospitalar

Allan da Silva Braga, Juliano Lopes de Oliveira

Instituto de Informática – Universidade Federal de Goiás (UFG)
74001-970 – Goiânia – GO – Brasil

{allanbraga, juliano}@inf.ufg.br

Abstract. *This article presents a case study on the implementation of a Service Desk function in a large public hospital with the support of the GLPI free software. The management of IT service processes were applied to the Service Desk function in accordance with the proposals of the ITIL framework. The scenario considered in the case study was that of a public hospital emergency unit, recently opened, with 24-hour service and a staff of about 2000 employees. During the deployment process difficulties were observed and specific measures implemented to enable the Service Desk operation in the context analyzed. The study provides evidence that free software can adequately meet the demands and contribute significantly to the improvement of the service requests flow and decision making by IT in a large hospital.*

Resumo. *Este artigo apresenta um estudo de caso realizado em um grande hospital público acerca da implantação de uma Central de Serviços (Service Desk) com o apoio do software livre GLPI. Os processos de gerenciamento de serviços de TI foram aplicados na função Service Desk de acordo com as propostas do framework ITIL. O cenário analisado no estudo de caso foi a unidade pública hospitalar de urgência, recém-inaugurada, com atendimento 24 horas e um quadro de cerca de 2000 funcionários. Durante o processo de implantação foram observadas dificuldades e implementadas medidas específicas para viabilizar a operação da Central de Serviços no contexto analisado. O estudo realizado provê indícios de que software livre pode atender adequadamente as demandas e contribuir significativamente para o aprimoramento do fluxo de atendimento de solicitações e tomada de decisão pelo setor de TI de uma unidade hospitalar de grande porte.*

1. Introdução

Nos últimos anos a gestão hospitalar vem sofrendo importantes transformações e significativos avanços, que auxiliam na condução das mais variadas atividades que envolvem esse ambiente considerado muito complexo [Mettler and Raptis 2012]. Nesse contexto, a Tecnologia da Informação (TI) tornou-se fundamental tanto para o processo operacional quanto para o processo gerencial, agregando controle financeiro, segurança, agilidade e apoio às decisões sobre todas as ações que ocorrem no ambiente hospitalar.

No Brasil, vários hospitais passam por dificuldades financeiras, possuem estruturas físicas e tecnológicas precárias, e carecem de profissionais qualificados. Essa situação dificilmente será resolvida sem que cada organização hospitalar tenha conhecimento adequado da sua própria realidade, com a disponibilidade de indicadores e informações confiáveis sobre a instituição, de forma a permitir que os administradores tomem decisões com base em dados necessários e apropriados. Nesse sentido, o apoio da TI é fundamental para melhorar a complexa gestão das organizações hospitalares, buscando assegurar a qualidade e a segurança das informações e a eficiência dos fluxos de atendimento nestas organizações.

O objetivo principal deste artigo é apresentar e discutir um estudo de caso sobre a implantação de recursos de TI para otimizar o trabalho em um ambiente hospitalar. O objetivo do estudo foi o de avaliar se a melhoria do Gerenciamento de Serviços de TI teria reflexos para usuários e clientes da unidade hospitalar. O estudo de caso envolveu a implantação da função Central de Serviços (ou *Service Desk*) em um hospital público de grande porte, tendo como base o software livre GLPI *(Gestionnaire Libre de Parc Informatique)* [Indepnet 2016]. Os processos e conceitos da Central de Serviços foram aplicados com base nas propostas do framework ITIL [Axelos 2011].

O restante desse artigo apresenta os principais aspectos do estudo de caso realizado e está organizado da seguinte forma. A Seção 2 apresenta a fundamentação teórica essencial para a compreensão do estudo. As seções 3, 4 e 5 discutem o estudo de caso. Finalmente, a Seção 6 apresenta conclusões do estudo realizado e aponta direções para trabalhos futuros.

2. Central de Serviços e Gestão de Serviços de TI

O estudo de caso descrito no presente artigo foi conduzido com base nas propostas de gestão de serviços de TI apresentadas na ITIL v3 [Axelos 2011]. ITIL é uma biblioteca de boas práticas para gerenciamento de serviços de TI que tem sido utilizada com sucesso por diversas organizações, públicas e privadas, com o objetivo de melhorar a qualidade dos seus serviços de TI, colocando-os em conformidade com os objetivos da organização.

As propostas da ITIL v3 (versão atual) são estruturadas em torno do ciclo de vida dos serviços dentro de uma organização que presta suporte de TI (para a própria organização ou para organizações externas). Esse ciclo de vida contempla:

- Estratégia do Serviço (*Service Strategy*): define requisitos do negócio e sua relação com os serviços de TI, estabelecendo o valor da TI para a organização cliente.
- Projeto de Serviço (*Service Design*): desenha a solução a ser adotada em cada serviço ofertado pela TI, de forma a atender as necessidades da organização.
- Transição de Serviço (*Service Transition*): estabelece a forma como serviços são entregues aos clientes e como ocorre o gerenciamento de mudanças nesses serviços.

- Operação do Serviço (*Service Operation*): especifica ações necessárias para garantir a satisfação dos compromissos assumidos em cada serviço, com base em um Acordo de Nível de Serviço (ou *Service Level Agreement* – SLA).
- Melhoria Contínua do Serviço (*Continual Service Improvement*): determina formas de avaliar o desempenho dos serviços visando a constante melhoria dos serviços e o seu alinhamento aos objetivos da organização.

Embora o estudo de caso descrito no presente artigo tenha sido conduzido com base nesse ciclo de vida de serviços de TI, o foco das discussões deste artigo é a fase de Operação do Serviço, pois é nela que o valor do serviço é entregue aos seus usuários e as demandas de TI são recebidas e atendidas pela organização fornecedora dos serviços.

A Operação de Serviço é critica dentro do ciclo de vida do serviço porque erros na condução, controle e gestão das atividades do cotidiano operacional podem comprometer a disponibilidade e a eficiência do serviço, mesmo que ele tenha sido bem planejado e que sua implementação tenha ocorrido com sucesso [Fernandes e Abreu 2008]. A Operação de Serviços contempla os seguintes processos [ITSMF 2007]:

- Gerenciamento de Evento: monitora todos os eventos que ocorrem na infraestrutura de TI e que podem indicar que algo não está funcionando como deveria, levando ao registro de um incidente.
- Gerenciamento de Incidente: restaura a operação normal de um serviço no menor tempo possível, no caso de ocorrência de um incidente.
- Gerenciamento de Problema: minimiza os impactos adversos de incidentes e problemas para o negócio, quando causado por falhas na infraestrutura de TI.
- Cumprimento de Requisição: trata as requisições dos usuários que foram geradas por uma solicitação de serviços de TI.
- Gerenciamento de Acesso: Controla o acesso de usuários aos serviços para manter a confidencialidade, disponibilidade e integridade dos dados.

A Central de Serviços (ou *Service Desk*) é uma função dentro da organização de TI que tem como propósito ser o único ponto de contato entre essa organização e os usuários e clientes de TI. Por ser a única interface operacional entre a fornecedora de serviços de TI e os seus usuários, a Central de Serviços é diretamente responsável pela percepção que os clientes de TI têm da organização fornecedora de serviços e, consequentemente, pela satisfação que esses clientes demonstram com relação aos serviços que recebem.

Quando corretamente implementada, a Central de Serviços funciona como uma entidade independente da organização fornecedora de serviços, assegurando que os clientes de TI recebam daquela organização as melhores práticas em relação à prestação de serviços de TI. Portanto, a Central de Serviços tem um papel estratégico para a prestação de serviços de TI e deve cumprir os seguintes objetivos [Bon 2005]:

- Funcionar como o ponto único de contato entre os usuários e a área de TI. A Central de serviços funciona como o 1º nível de suporte aos usuários.

- Restaurar os serviços sempre que possível. A equipe de suporte deve estar equipada com ferramentas e informações, tais como erros conhecidos e bases de conhecimento, para que possa oferecer soluções de contorno o mais rápido possível.
- Prover suporte com qualidade para atender os objetivos de negócio da organização cliente. A equipe deve estar bem treinada para ter conhecimento de todos os serviços fornecidos e entender o impacto que eles têm para o negócio.
- Gerenciar todos os incidentes até o seu encerramento. A Central de Serviço é responsável pelo processo de Gerenciamento de Incidentes, ou seja, pelo tratamento de todos os incidentes. Também é responsabilidade da Central de Serviços monitorar o cumprimento dos acordos estabelecidos em ANS (Acordo de Nível de Serviço).
- Dar suporte a mudanças, fornecendo comunicação aos usuários e mantendo-os informados sobre todos os eventos envolvidos nas mudanças de serviços.
- Aumentar a satisfação do usuário, provendo suporte com qualidade, mantendo prontidão para o atendimento, e buscando solucionar incidentes de forma ágil de modo a maximizar a disponibilidade do serviço.

Para a implantação da Central de Serviços no presente estudo de caso foi configurada a ferramenta GLPI (*Gestionnaire Libre de Parc Informatique*), software livre de código aberto liberado sob a licença GNU/GPL v2.0 e construído para apoiar o gerenciamento de ativos de TI, inclusive a solução de Central de Serviço.

O GLPI é uma aplicação Web desenvolvida em PHP que realiza o inventário de componentes de um parque de TI (hardware e software), além de oferecer funcionalidades para gestão de suporte ao usuário e de ativos de TI, conforme sintetiza a Tabela 1.

Tabela 1. Funcionalidades e características do software GLPI

Item	Funcionalidades e Características
Inventário	Inventário de computadores, periféricos, impressoras de rede e componentes associados através de uma interface com o inventário OCS ou FusionInventory. Gestão administrativa, contratos, documentos relacionados com componentes de inventário, estado de equipamento e de reservas. Gestão de pedidos de assistência de todos os tipos de inventário de equipamentos. Gestão da informação financeira (compra, garantia e extensão, amortização).
Central de Serviços	Gerenciamento de problemas em vários ambientes via criação de ticket, planejamento, alocação e gestão de bilhetes. Comentário de

	solicitação/requisição. Rastreamento de e-mail e gestão de demandas de intervenção. Gerenciamento de problemas, projeto e mudanças. Histórico de intervenções. Pesquisa de satisfação. Escalonamento de demandas. Gerenciamento de conteúdo por metas (perfis, grupos, etc.). Sistema básico de gestão de base de dados de conhecimento. FAQ. Interface de usuário (calendário).
Relatórios e Estatísticas	Relatórios em vários formatos (png, svg, csv, etc.). Estatísticas globais. Categorias de estatísticas (por técnico, hardware, usuário, categoria, prioridade, localização, etc.). Relatórios de dispositivos (tipo de dispositivo, contrato associado, informações comerciais, etc.).

3. Estudo de Caso – Cenário Inicial

Esta seção descreve a situação do setor de TI da unidade hospitalar antes da implantação da Central de Serviços. A descrição é feita no tempo presente para manter a maior fidedignidade possível com relação às observações feitas pelos pesquisadores naquela oportunidade.

3.1. Estrutura da equipe de TI

O setor de TI da unidade hospitalar é composto por vinte e três colaboradores, sendo um gestor, quatro analistas e dezoito técnicos em Informática. Essa equipe é responsável pelo suporte ao usuário, pela manutenção da infraestrutura de TI e pelo funcionamento das atividades que envolvem TI.

A equipe mantém sob sua supervisão o funcionamento de aproximadamente 700 computadores, além de dispositivos móveis (palms, tablets, etc.), impressoras, scanners, equipamentos de projeção e videoconferência, além de equipamentos de TI voltados especificamente para propósitos hospitalares que se comunicam em rede.

O setor de TI realiza atendimento aos usuários 24 horas por dia e atualmente é estruturado em três áreas:

- Suporte: Responsável pela manutenção da rede física cabeada, suporte e atendimento aos usuários, manutenção de hardware e software, e manutenção preventiva e corretiva dos ativos de TI.
- Infraestrutura: Responsável por todo o funcionamento lógico da rede de TI e pelos equipamentos servidores.
- Estatística/desenvolvimento: Responsável pelo desenvolvimento de relatórios e indicadores de gestão da área de TI, e pela manutenção da intranet institucional.

Para atendimento aos usuários essas áreas trabalham de formas distintas. A área de suporte é acionada por um Sistema de Atendimento de Chamados de Tecnologia da Informação (SATI) e também por telefone. O atendimento pode ser realizado via acesso remoto ou presencial, dependendo da gravidade do chamado. Os atendimentos

solicitados às áreas de Infraestrutura e Estatística são realizados somente via SATI ou por e-mail.

3.2. Atendimento de Chamados de Tecnologia da Informação

O sistema SATI utilizado pelo hospital apresenta uma estrutura simples e direta, como pode ser observado na Figura 1. Trata-se de uma aplicação desenvolvida pela própria TI do hospital, integrado à sua intranet.

O sistema SATI possibilita que o usuário acione o setor de TI através de uma interface simples, registrando seu pedido de atendimento. Os pedidos de atendimento são enfileirados por ordem de solicitação. Por ser um setor que trabalha em todos os turnos (diurno e noturno) de forma ininterrupta, os atendimentos nem sempre são concluídos pelo mesmo técnico. Por exemplo, uma ocorrência aberta pela equipe de técnicos do horário diurno pode ser encerrada por uma equipe de técnicos do turno noturno.

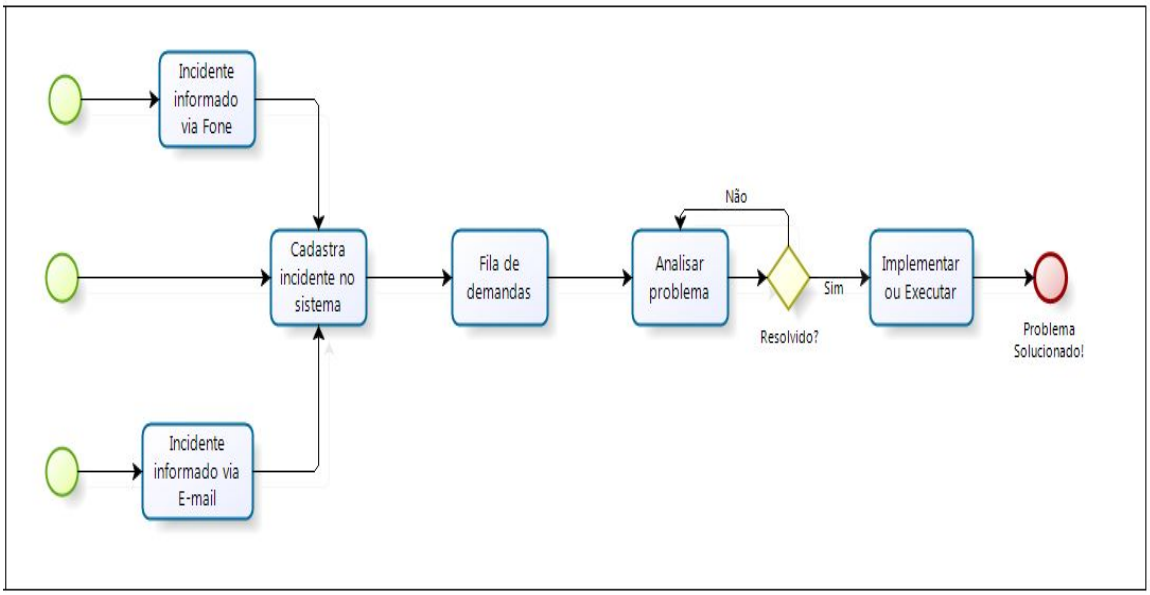


Figura 1 - Fluxo de atendimento no sistema SATI

Não existe registro do técnico responsável diretamente pelo atendimento da ocorrência; somente é feito o registro de sua participação em determinada etapa, o que dificulta o controle estatístico de atendimento por técnico.

Os seguintes pontos positivos e negativos podem ser destacados em relação ao processo de gestão de serviços de TI apoiado pelo SATI:

- Pontos positivos:
  - A interface do software é simples e objetiva para o usuário.
  - Não há necessidade de imprimir formulários de atendimento para o técnico.
  - O sistema apresenta poucas ocorrências de falhas em seu funcionamento.
- Pontos negativos:

- Há somente uma opção de login, na qual o colaborador deve informar o número de matrícula. Porém, funcionários terceirizados, que também dependem do sistema, não possuem este pré-requisito.
- Há poucas opções de categorias para abertura dos chamados.
- Serviços não aceitos pelo SATI são solicitados por telefone ou e-mail.
- Não existe classificação de prioridades das solicitações.
- Não há níveis de atendimento.
- O controle estatístico de atendimentos/chamados é limitado.
- Não há prazos estipulados para controle, monitoramento e atendimento.
- Não existe uma base de conhecimento que possibilite agilidade na solução de um incidente.

4. Estudo de Caso – Cenário Após Implantação da Central de Serviços

Com o apoio do software GLPI foi proposta a criação de níveis de serviços para atendimento e suporte aos usuários. Para atendimento em nível 1 foi necessário manter dois técnicos com atribuições relacionadas à Central de Serviços. Eles realizam a triagem das entradas geradas e procuram atender remotamente essas solicitações.

Caso não seja possível resolver o incidente, o técnico avança a demanda para o nível 2 de suporte, que é atribuído a um técnico de plantão que se desloca ao local para avaliar e intervir na solução do incidente, de acordo com a prioridade classificada em SLA.

A Figura 2 ilustra o novo fluxo de atendimento baseado no GLPI, incluindo etapas que ainda não foram completamente finalizadas na implantação da Central de Serviços, como Gerenciamento de Nível de Serviço com os devidos Acordos de Nível de Serviço (SLAs) e a base de conhecimento recomendada pela ITIL.

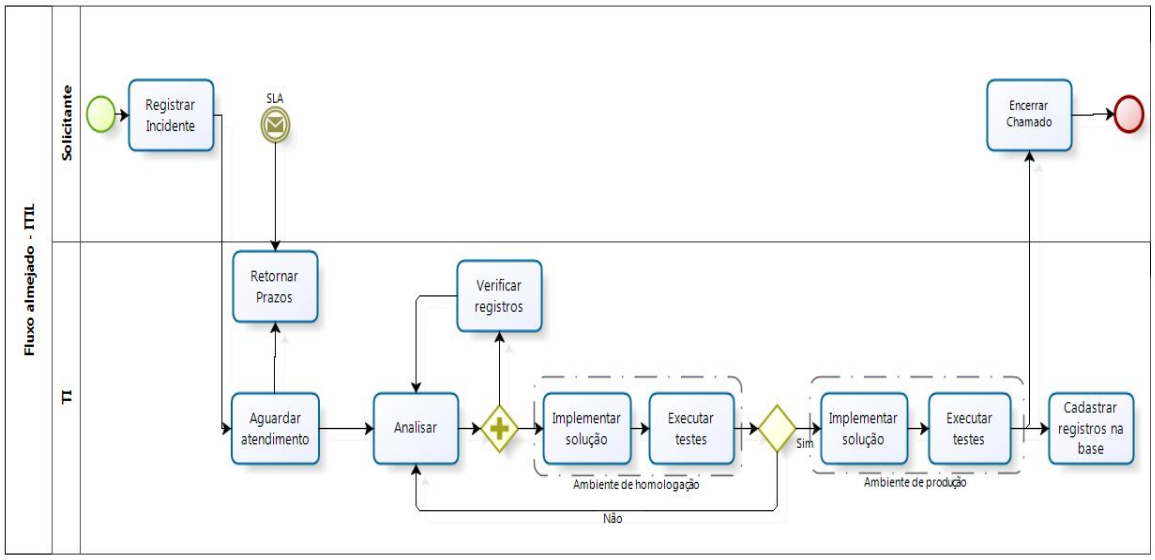


Figura 2 - Fluxo de atendimento após a implantação do GLPI

Com o fluxo implementado na nova ferramenta é possível contabilizar nos prazos do serviço o período de retorno de informações dos usuários, de forma que os Acordos de Níveis de Serviço (SLAs) possam ser cumpridos quando há interferência do usuário no fluxo do processo. A implantação de SLA dos diversos serviços está em fase de negociação com os gestores responsáveis pelos diversos setores do hospital.

Enquanto os SLAs não são homologados, os prazos estão sendo definidos pelo Gestor da área de TI. Para cada prazo definido é realizado um monitoramento do tempo de resolução esperado dos chamados com relação ao tempo real de entrega do serviço para fornecer insumos para estabelecer SLAs mais realistas. Após a homologação dos SLAs pelos gestores das áreas clientes dos serviços será possível priorizar os chamados, levando em consideração a sua criticidade e impacto para o negócio, conforme especificado no respectivo SLA.

Outro aspecto que contribui para agilizar o atendimento dos chamados pelo setor de TI é a Base de Conhecimento disponibilizada no GLPI. Os chamados frequentes começam a apresentar tempo de resolução menor, dado a uniformização das soluções para incidentes recorrentes. Esta Base de Conhecimento é acessada no próprio GLPI.

4.1. Principais Dificuldades Encontradas na Implantação da Central de Serviços

A realização do treinamento dos envolvidos foi uma das tarefas mais complexas na implantação realizada neste estudo de caso, pois o expediente no hospital contínuo, 24 horas por dia, 7 dias por semana (24/7).

O suporte ao colaborador também acompanha esse expediente, de forma que além de preparar bem a equipe da TI, foi necessária a formação de diversas turmas (manhã, tarde e noite) para capacitação dos usuários do corpo clínico e da área administrativa do hospital no uso do GLPI.

Nem todos os funcionários possuíam *login* próprio (muitos compartilhavam um mesmo login) para acesso ao sistema. Como a ferramenta GLPI foi configurada para utilizar essa forma de autenticação, foi necessária uma força tarefa da equipe de desenvolvimento da área de TI para montar um script que incluísse todos os usuários no *Active Directory* (AD) da organização.

Após o cronograma estabelecido pela TI e validado pela direção da unidade, muitos colaboradores alegaram que ainda não haviam recebido a capacitação inicial e apresentaram resistência para abertura de chamados na nova ferramenta. Com isso, houve a necessidade de abertura de turmas “extras” para o treinamento na nova ferramenta de Central de Serviços.

Outra dificuldade encontrada na implantação do novo fluxo de gestão de serviços de TI foi relacionada à alocação de técnicos na Central de Serviços. Como o funcionamento do setor de TI acompanha o funcionamento contínuo do hospital, não foi possível contar com colaboradores responsáveis exclusivamente pela Central de Serviços em todos os horários, principalmente durante o final de semana, quando a equipe se encontra reduzida aos técnicos de plantão. Logo, o fluxo ideal do processo definido passou a vigorar principalmente de segunda à sexta em horário comercial (a ferramenta GLPI, possibilitou essa tomada de decisão para definição dos horários em que o processo completo pode ser executado).

Logo após o início das atividades do novo processo, mesmo depois de boa parte dos colaboradores da organização já terem sido treinados, surgiu uma resistência ao uso da nova metodologia de trabalho da TI. O fato de estarem todos (usuários e técnicos de TI) em um mesmo espaço físico (do hospital) contribuiu para isso. A proximidade física provia uma oportunidade para que o usuário (funcionário) saísse de sua mesa de trabalho e fosse até o setor de TI pessoalmente, com a finalidade de buscar solução para seu problema. Diante dessas situações, os próprios técnicos de TI foram reticentes quanto à viabilidade dos novos processos serem realmente executados e respeitados por todos os envolvidos.

Segundo [Magalhães e Pinheiro 2007], um processo de mudança compartilha diversos momentos distintos. Para obtenção do sucesso na execução de uma mudança seria necessário que os envolvidos participem ativamente em:

- Reconhecer a necessidade da mudança.
- Conhecer a "visão da mudança".
- Reconhecer as condições limitantes.
- Selecionar o método a ser utilizado na mudança.
- Implementar e avaliar o método utilizado para introduzir a mudança.

Em uma organização de grande porte, com cerca de dois mil usuários, não é viável garantir todas essas condições que favorecem uma mudança mais tranquila em processos de TI. Dessa forma, as medidas que precisaram ser tomadas para minimizar o risco de insucesso do novo processo foram centradas em reforçar a ideia de que o processo é importante para a organização como um todo.

Nesse sentido, enviou-se, via e-mail, comunicados a todos colaboradores, informando sobre o apoio da direção da instituição aos novos serviços e aos novos processos adotados pela área de TI. Além disto, foi realizada uma reunião da Diretoria Administrativa com todos os supervisores, com o intuito de reforçar o apoio institucional à nova rotina de trabalho instaurada.

5. Discussão dos Resultados

Após cerca de um mês de implantação da Central de Serviços na organização, foram observados os resultados discutidos a seguir.

Como a equipe de TI trabalha em escala de revezamento (inclusive em finais de semana e feriados), e considerando os gráficos apresentados na Figura 3, concluiu-se que as demandas possuem volume muito menor aos sábados e domingos em relação ao restante da semana, e que os horários com maior fluxo de demandas ocorrem entre 7 horas da manhã e 8 horas da noite. Isso possibilitou o remanejamento dos colaboradores de maneira que, nos dias e períodos de pico no atendimento, houvesse um maior número de técnicos para prestação do suporte.

Os dados coletados referem-se ao período de 15 de setembro à 12 de outubro de 2016, correspondendo ao início da implantação até a data de conclusão deste trabalho.

Após o remanejamento dos colaboradores para os dias e horários com maior fluxo de atendimento, iniciou-se a configuração da base de conhecimento disponível na

ferramenta GLPI, o que possibilitou uma diminuição do tempo de resolução de chamados. Na última medição realizada, cerca de 80% dos chamados foram solucionados em menos de um dia, conforme mostra a Figura 4.

Diante da melhor distribuição dos profissionais da TI e do aprimoramento da agilidade na resolução dos atendimentos, foi possível constatar um elevado índice de satisfação dos usuários (índice geral de 92% de satisfação, no período avaliado, conforme figuras 5 e 6).

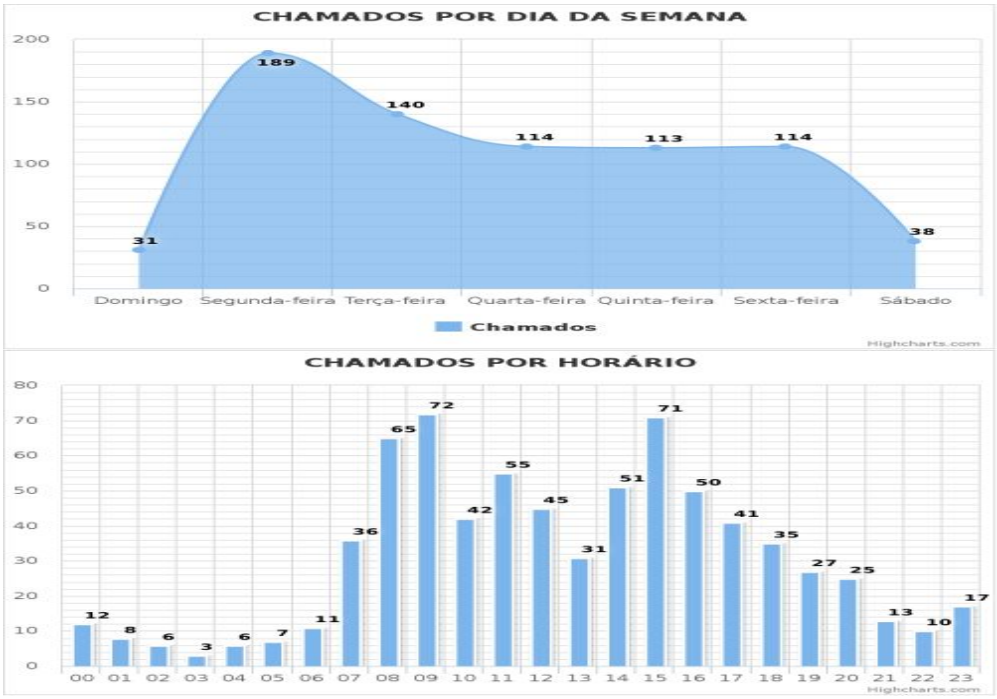


Figura 3 - Quantidade de chamados por dia x horário

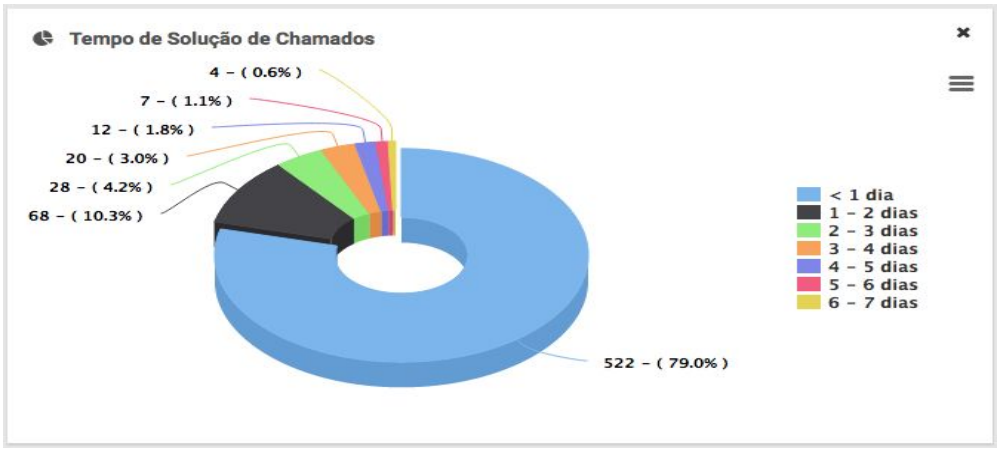


Figura 4 - Tempo de Resolução dos Chamados

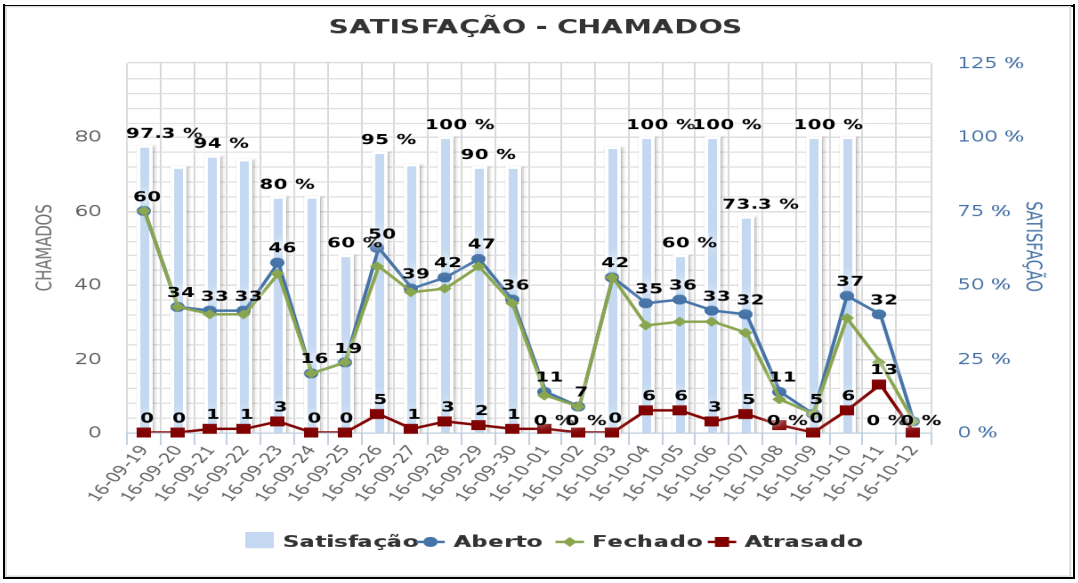


Figura 5 – Pesquisa de Satisfação dos Clientes

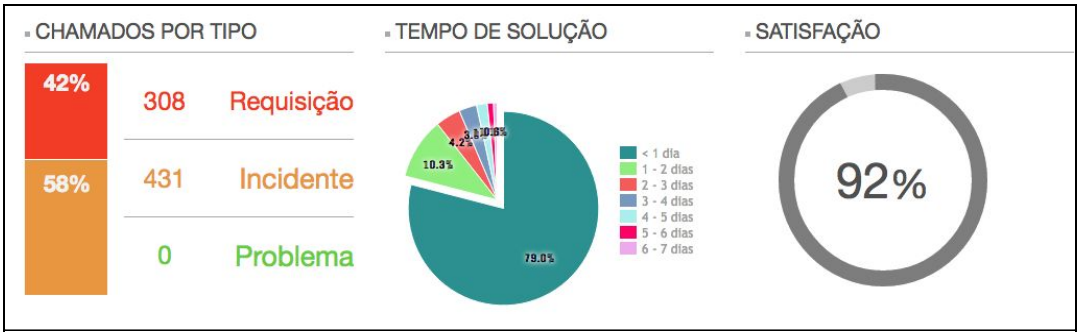


Figura 6 - Resumo dos Indicadores da Central de Serviços

6. Conclusões

Este artigo discutiu os resultados de um estudo de caso de implementação de uma Central de Serviços em uma grande unidade hospitalar pública. Nesse estudo, foi realizada uma análise da situação inicial das ferramentas de controle e gestão do setor de TI do hospital. Após a implementação da Central de Serviços foi feita a análise das mesmas situações. Os resultados mostram que o uso das ferramentas de gerenciamento de ativos integradas pelo GLPI representou um ganho para a gestão de serviços de TI do hospital. Vale ressaltar que a ferramenta GLPI foi implantada e está sendo utilizada, porém o sistema e os processos ainda estão em fase de adaptação.

Os conceitos da Operação de Serviços da ITIL, voltados à entrega de níveis de serviços com a qualidade e a rapidez que o negócio exige, necessitam das organizações uma conduta de tratamento de incidentes eficiente, que seja capaz de realizar monitoramento dos serviços e atuar de forma consistente para a solução de chamados. Foi possível identificar no presente estudo de caso alguns pontos relevantes da gestão de incidentes conduzida pelo setor de TI de forma a propor um novo modelo que

aprimorasse a prestação de serviços do setor. Verificou-se que para a implantação das melhores práticas da ITIL, em específico da função de Central de Serviços, o setor de TI deveria estar estruturado e alinhado aos interesses da instituição.

Através de análise dos procedimentos e ferramentas utilizadas, chegou-se à conclusão que seria preciso organizar o fluxo e os processos de trabalho executados. Tendo em vista o custo para aquisição de uma ferramenta proprietária ou mesmo para o desenvolvimento/melhoria da ferramenta utilizada anteriormente pelo setor de TI, optou-se pela implantação e uso da ferramenta de gerenciamento de ativos GLPI, um software livre que permite uma unificação das informações sobre máquinas, periféricos e equipamentos de rede, além de adaptações às necessidades específicas do hospital.

Os serviços antes oferecidos operavam com o uso de ferramentas desenvolvidas localmente, limitadas e com poucas opções para o acompanhamento dos chamados. A geração de métricas e relatórios de atendimento eram desenvolvidas em outra ferramenta particular, portanto, antes do emprego do GLPI, não era possível monitorar e acompanhar acordos de nível de serviços ou saber o status de um atendimento com seu devido registro e seu histórico de movimentações de entradas e saídas.

Com a implantação e uso da ferramenta na unidade hospitalar foi possível implementar o controle dos atendimentos realizados, organizar a equipe de TI e o controle do tempo para realização de tarefas, bem como melhorar o fluxo de trabalho. Os resultados mostraram importantes melhorias em todo o processo, principalmente na tomada de decisão pelo setor de TI quanto a sobrecarga de trabalho que incidia sobre os técnicos em determinados horários e dias da semana. Com isso, foi possível preencher lacunas em horários de picos e reduzir a ociosidade aos finais de semana e feriados com a redução da equipe nestas ocasiões.

A implantação segue em curso e estamos aperfeiçoando a etapa de Operação de Serviços, com a implementação de processos que estão em adaptação, como o de Gerenciamento de Eventos e o Gerenciamento de Problemas. Isso possibilitará a prática da melhoria contínua dos serviços, como propõe a ITIL.

Referências Bibliográficas

Axelos. ITIL - Information Technology Infrastructure Library - Service Operation. London, 2011.

Bon, Jan von. Foundations of IT Service Management, based on ITIL. Lunteren - Holanda: Van Haren Publishing, 2005.

Cohen, Roberto. Implantação de Help Desk e Service Desk. 1 ed. Ed. Novatec, 2008.

Costa, André Cypriano Monteiro. Modelagem do domínio do processo de gerenciamento de nível de serviço do padrão ITIL uma abordagem usando ontologias de fundamentação e sua aplicação na plataforma infraware. 2008.

Fernandes, Aguinaldo Aragon; Abreu, Vladimir Ferraz. Implantando a governança de TI: da estratégia à gestão dos processos e serviços. 2. ed. Rio de Janeiro: Brasport, 2008.

Freitas, Marcos André dos Santos. Fundamentos do Gerenciamento de Serviços de TI. Editora Brasport, 2010.

Indepnet. GLPI Project: Gestion Libre de Parc Informatique. http://www.glpi-project.org. 2016.

ITSMF - IT Service Management Forum. An Introductory Overview of ITIL V3. 2007. http://www.best-management-practice.com/gempdf/itSMF_An_Introductory_Overview_of_ITIL_V3.pdf.

Magalhães, I.L.; Pinheiro, W.B. Gerenciamento de Serviços de TI na Prática. Ed. Novatec, 2007.

Mettler T, Raptis D. A. What constitutes the field of health information systems? Fostering a systematic framework and research agenda. Health Informatics Journal. 18 (2): 147–156, 2012.

PMI, Project Management Institute. Um Guia do Conhecimento em Gerenciamento de Projetos (Guia PMBOK). 4. ed. Pensilvânia: Global Standard, 2008.

Yin, R. K. Case Study Research: Design and Methods (3rd Edition). CA: Sage, 2002.

Aplicação do Padrão ISO/IEEE 11073 no contexto da Assistência Domiciliar à Saúde

Douglas Battisti, Sergio Teixeira de Carvalho

Instituto de Informática (INF) – Universidade Federal de Goiás (UFG)
Câmpus Samambaia – Goiânia – GO – Brasil

dougbattisti.5@gmail.com, sergio@inf.ufg.br

Abstract. *The home health care is recognized as one of the mechanisms for reducing the number of hospital admissions, as well as a way to provide welfare to the patient being treated at home, especially elderly people. Through ubiquitous computing techniques, involving physiological sensors and medical devices used at the home environment, the patient can be monitored continuously and remotely anywhere in the house. For effective monitoring, the solution needs to be interoperable on the variety of medical devices and physiological sensors. In this sense, this paper presents a software solution that can provide physiological data from sensors and medical devices, adopting the communication standard ISO/IEEE 11073.*

Resumo. *A assistência domiciliar à saúde é reconhecida como um dos mecanismos para a redução do número de internações hospitalares, como também uma forma de proporcionar bem-estar ao paciente tratado em casa, em especial pacientes idosos. Por meio de técnicas de computação ubíqua, envolvendo sensores fisiológicos e dispositivos médicos utilizados no ambiente domiciliar, o paciente pode ser acompanhado continuamente e de forma remota, em qualquer parte da casa que esteja. Para um efetivo acompanhamento, a solução necessita ser interoperável diante da variedade de dispositivos médicos e sensores fisiológicos. Nesse sentido, este artigo apresenta uma solução de software que seja capaz de disponibilizar dados de sensores fisiológicos e dispositivos médicos, adotando o padrão de comunicação ISO/IEEE 11073.*

1. Introdução

Devido aos avanços científicos e tecnológicos das últimas décadas, houve um expressivo aumento na expectativa de vida das pessoas, e, consequentemente, do número de pessoas dependentes de serviços prestados pela área da saúde. Em se tratando especialmente de idosos e pacientes com doenças crônicas (e.g., hipertensão), há ainda a necessidade de um acompanhamento contínuo do seu estado de saúde. Nesse sentido, a assistência domiciliar à saúde viabilizada pela aplicação de técnicas de computação ubíqua, tem sido investigada como uma alternativa para lidar com esses problemas [Wood et al. 2008, Sztajnberg et al. 2009, He et al. 2013, Kim and Chung 2014, Theoharidou et al. 2014, Brauner et al. 2015, Karthick et al. 2016].

Este artigo busca soluções em termos de implementação de *software* voltadas para sensores fisiológicos e dispositivos médicos, como, por exemplo, termômetros,

balanças, medidores de pressão arterial e glicosímetros, no contexto da assistência domiciliar à saúde. A Continua Health Alliance, uma organização sem fins lucrativos formada por um grupo de companhias tanto da área de saúde quanto da área tecnológica, fornece uma estruturação e diretrizes de interoperabilidade relacionadas a tais sensores e dispositivos, com base na família de padrões ISO/IEEE 11073 Personal Health Data [Piniewski et al. 2010].

Os objetivos deste trabalho são investigar o Padrão ISO/IEEE 11073 especificamente voltado a dois tipos de sensores – medidor de pressão arterial e balança – e implementar uma solução de *software* que capture, manipule e trate dados desses tipos de sensores. Para fins de experimentação, foi utilizado o Plano de Cuidados Ubíquo, uma solução computacional para assistência domiciliar à saúde, em desenvolvimento no Laboratório de Informática em Saúde (LabIS) do Instituto de Informática da Universidade Federal de Goiás [Germano et al. 2016]. O LabIS é parte do Laboratório de Redes de Computadores e Sistemas Distribuídos (LABORA) do Instituto de Informática.

Este artigo está organizado em mais 5 seções. A Seção 2 apresenta conceitos de Assistência Domiciliar à Saúde e Computação Ubíqua, enquanto que a Seção 3 detalha o padrão ISO/IEEE 11073 envolvendo seus principais componentes e a sua arquitetura. A Seção 4, por sua vez, apresenta a implementação da solução de *software*, com destaque para a metodologia utilizada. Os resultados estão na Seção 5, na qual é apresentado o protótipo da solução. Por fim, a Seção 6 traz as conclusões e as possibilidades de trabalhos futuros.

2. Assitência Domiciliar à Saúde Aliada à Computação Ubíqua

Em 1991, Mark Weiser publicou o artigo "*The computer for the 21st century*" [Weiser 1991] , onde, pela primeira vez, o termo Computação Ubíqua (*Ubiquitous Computing*) foi usado. Nesse trabalho foram previstas algumas mudanças que ocorreriam na computação nas décadas seguintes, tendo em vista seu crescimento e a miniaturização dos dispositivos computacionais. Weiser teorizou que o foco do usuário estaria voltado para a execução de tarefas, e não para as ferramentas utilizadas, empregando-se a computação sem perceber ou necessitar de conhecimentos a respeito da máquina e do *software* envolvido. Hoje, a invisibilidade da computação sob a perspectiva do usuário viabiliza o uso de tecnologias computacionais como forma de auxiliar o crescimento de várias áreas do conhecimento, como, por exemplo, a saúde.

A aplicação da computação ubíqua no domínio da assistência domiciliar à saúde pode ser vista como um complemento ao tratamento convencional [Copetti et al. 2008, Sztajnberg et al. 2009]. Essa união pode trazer para o paciente, sendo a maioria idosos, a facilidade de disponibilizar seus dados fisiológicos para aqueles que se interessam, como, por exemplo, seu médico ou mesmo cuidador. A facilidade estaria no mínimo de conhecimento que o usuário precisaria para conseguir compartilhar seus dados fisiológicos. E por se tratar de um tratamento em casa a presença da computação deve ser invisível, não interferindo na rotina do paciente.

3. O Padrão ISO/IEEE 11073 Personal Health Data

A interoperabilidade entre sistemas de saúde e dispositivos médicos pessoais é o problema atacado pela família de padrões ISO/IEEE 11073 PHD (X73). O padrão define um

protocolo de troca de dados eficiente, bem como os modelos de dados necessários para a comunicação entre dispositivos.

3.1. Agent e Manager

O padrão X73 define que na comunicação existem dois elementos denominados por Agent e Manager. O Agent é o lado da comunicação que produz os dados, podendo ser exemplificado como uma balança, um glicosímetro, ou um medidor de pressão arterial. O Agent, por ser um dispositivo eletrônico pessoal e frequentemente vestível, deve possuir algumas características, como pouco gasto de energia, para estender a autonomia do aparelho, e CPU limitada, para diminuir o tamanho e o peso do dispositivo. Portanto, o protocolo de comunicação precisa ser leve, evitando mensagens longas, e ser eficiente em termos de *overhead*, largura de banda e uso da CPU [Yao and Warren 2005].

O Manager, por sua vez, é o lado da comunicação que recebe as mensagens, além de interpretá-las e disponibilizá-las ao sistema de assistência domiciliar à saúde. Pode ser exemplificado como um *notebook, personal computer, smartphone, smartwatch*, *hub* de comunicação, etc. Conforme a interface de comunicação do Agent, é melhor usar um motor computacional específico, como *notebook* ou *smartphone* [Piniewski et al. 2010]. Por exemplo, um Agent pode ter somente comunicação *Bluetooth*, favorecendo o uso de smartphone pela mobilidade, já que na maioria das vezes ele acompanha o paciente. Além da troca de mensagens com o Agent, o Manager deve enviar os dados para o sistema. O envio dos dados para o sistema pode ocorrer de várias formas, como, por exemplo, via comunicação entre processos, XML, JSON e Sockets.

Um típico cenário de uso do X73 é ilustrado na Figura 1. O dispositivo médico pessoal (Agent), como medidor de pressão arterial, balança e glicosímetro, captura dados fisiológicos do sensor e os envia para o aparelho receptor (Manager), como *smartphone*, *notebook* ou *hub* de comunicação. O Manager então processa os dados e os disponibiliza para o sistema de Assistência Domiciliar à Saúde.

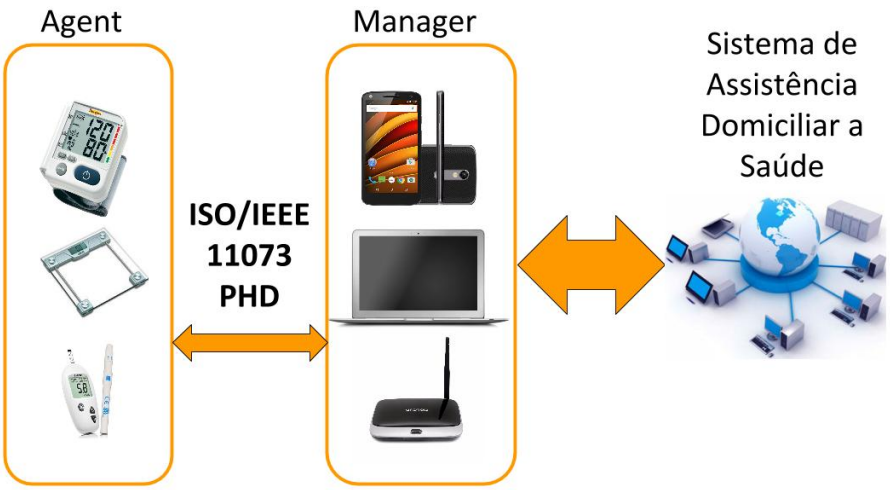


Figura 1. Cenário de uso da família X73. Adaptado de [Yu et al. 2013].

3.2. Arquitetura

A família X73 é organizada utilizando três grandes modelos: o modelo de domínio da informação (DIM), o modelo de serviço e o modelo de comunicação. No modelo de domínio da informação, um dispositivo médico é modelado como um objeto com múltiplos atributos. Esses atributos indicam opções de configuração, dados fisiológicos e outras funcionalidades particulares. O modelo de serviço, por sua vez, define o processo de acesso aos dados como *GET, SET, ACTION* e *Event-Report* entre um Agent e um Manager. O terceiro modelo, o de comunicação, descreve um protocolo de conexão ponto a ponto entre um Agent e um Manager usando a máquina de estados do Agent e a máquina de estados do Manager.

Subpadrão da família X73, o ISO/IEEE 11073-20601[1] é o coração da troca de mensagens, uma vez que define como múltiplos Agents podem se conectar com um Manager em uma comunicação ponto a ponto independente da camada de transporte. Um cenário que ilustra como Agent e Manager trocam dados é mostrado na Figura 2.

A troca de dados tem início com o Agent enviando para o Manager um *Association request* contendo o *system ID* da balança, o número de versão do protocolo e outras informações de configuração do dispositivo. Se o Manager reconhece o *system ID*, envia uma mensagem de aceitação – *Association response (accepted)*, fazendo com que ambos entrem no estado de *Operation*. O Agent, então, envia os dados fisiológicos para o Manager – *Confirmed Event Report (weight)*. O Manager recebe com sucesso os dados fisiológicos e responde com a confirmação – *Acknowledgement (OK)*. Finalmente, o Agent requisita o fim da associação e o Manager responde a essa requisição.

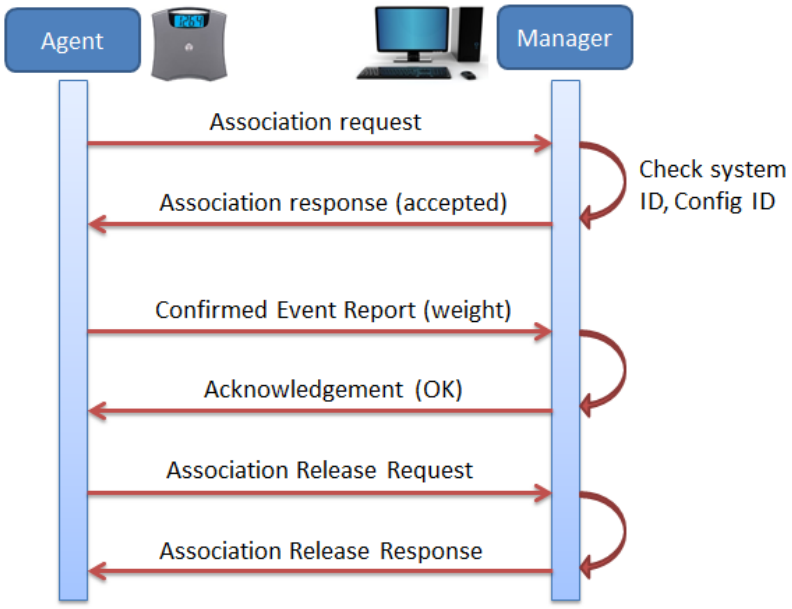


Figura 2. Cenário de troca de dados entre Agent e Manager. Adaptado de [Yu et al. 2013].

[1] https://standards.ieee.org/findstds/standard/11073-20601-2014.html

3.3. Continua Alliance

A complexidade e densidade dos documentos da família X73 é um dos pontos que restringem sua adoção [Martinez et al. 2007]. A Continua Health Alliance[2] é uma organização sem fins lucrativos formada por aproximadamente 240 empresas na área da saúde e tecnologia, como, por exemplo, Samsung, IBM, A&D Medical e Omron. Elas se dedicam a reduzir custos e melhorar os resultados, por meio do desenvolvimento de guias técnicos de interoperabilidade para sistemas de monitoramento e dispositivos médicos, que auxiliam no entendimento e na aplicação dos padrões. Além disso, a Continua certifica os produtos que estão exatamente enquadrados nas normas especificadas no padrão X73 [Piniewski et al. 2010].

4. Implementação

No sentido de implementar uma solução de *software* para capturar e manipular dados envolvendo o padrão ISO/IEEE 11073, foi implementado um módulo cujo objetivo é capturar e disponibilizar os dados fisiológicos obtidos por sensores médicos. Como estudo de caso, foram processados dados fisiológicos referentes à pressão arterial e peso. O módulo de *software* foi implementado para realizar:

- a obtenção dos dados fisiológicos por meio de dispositivos de medição e sensores;
- o tratamento e a disponibilização dos dados para uma camada superior de *software*, denominada UbiCare, visando a integração com o sistema de Plano de Cuidados Ubíquo;
- a integração com dispositivos Android, a fim de simular um dispositivo de medição;

A interoperabilidade foi introduzida a partir da adoção do padrão, utilizando a biblioteca de *software* aberta Antidote [Signove 2012], discutida na Subseção 4.3.

4.1. Plano de Cuidados Ubíquo

Um plano de cuidados refere-se ao conjunto de prescrições voltadas ao paciente, tais como as medições que o paciente deve realizar e com qual frequência, quais medicamentos e quando devem ser ingeridos, além de recomendações voltadas à dieta e exercícios físicos, dentre outras. [Haynes et al. 2002]

A implementação de um Plano de Cuidados Ubíquo tem como objetivo trocar a forma manual e complexa de elaboração tradicional do plano de cuidados por uma implementação de *software* em que o paciente em tratamento possa ser notificado por meio do uso de tecnologias que são comuns em seu cotidiano, como, por exemplo, *smartphones, tablets* e TVs. Além disso, essas tecnologias podem oferecer uma melhor interação tanto do paciente com o plano de cuidados, para efeitos de visualização das prescrições, quanto do cuidador com o mesmo, para a verificação do cumprimento delas. [Germano et al. 2015]

4.2. UbiCare

O UbiCare é uma solução arquitetural idealizada no contexto do grupo de pesquisa do Laboratório de Informática em Saúde com o propósito de realizar de forma transparente,

[2] www.continuaalliance.org

tanto para pacientes quanto para cuidadores e profissionais de saúde, operações no contexto da assistência domiciliar à saúde e do monitoramento de pacientes. Essas operações podem variar em função de uma série de fatores, como, por exemplo, o tipo de tratamento e o estado de saúde do paciente. De modo a coordenar as ações dessas aplicações, garantir a interoperabilidade, manter um baixo acoplamento e permitir uma reutilização de informações que são comuns entre elas, o modelo tem como base uma arquitetura orientada a serviços, do inglês, Service-Oriented Architecture (SOA) [Endrei et al. 2004].

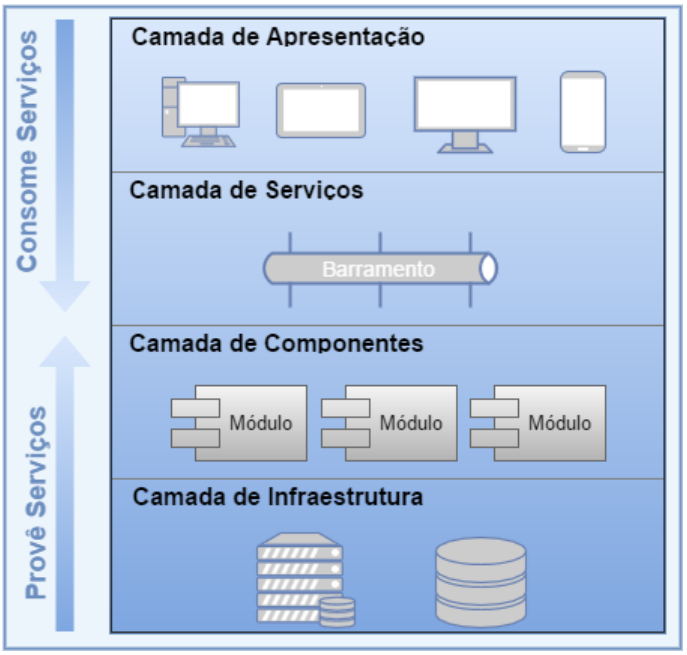


Figura 3. Camadas arquiteturais do UbiCare.

A Figura 3 mostra a representação das camadas arquiteturais do UbiCare. A camada superior, chamada apresentação, é onde são encontradas as aplicações ubíquas com suas interfaces de acesso pelos usuários. Abaixo está a camada de serviços que funciona como uma interface para o acesso dessas aplicações a recursos de *software* compartilhados entre elas. Esses recursos são representados na camada de componentes, que por sua vez, consomem recursos de mais baixo nível da camada de infraestrutura, como os disponibilizados por sistemas gerenciadores de banco de dados e sistemas operacionais.

O modelo Plano de Cuidados Ubíquo é sintetizado a partir dessa arquitetura, onde as aplicações que interagem diretamente com pacientes e envolvidos no tratamento encontram-se na Camada de Apresentação. Essas aplicações podem estar embarcadas em quaisquer dispositivos (e.g., *smartphones*) que possam se comunicar usando um barramento de serviços seguindo um determinado padrão de comunicação. Esse barramento pode ser um *web service* que disponibiliza uma *Application Programming Interface (API)* com um conjunto de componentes de *software* que implementam as funcionalidades de um plano de cuidados. Além disso, recursos mais avançados, no sentido de exceder as funcionalidades básicas de um plano de cuidados convencional, podem ser implementados e/ou gerenciados nesses componentes.

No contexto deste artigo, foi implementado um módulo na camada de componentes desta arquitetura, fazendo parte do modelo Plano de Cuidados Ubíquo. Para viabilizar

os testes, foi implementada também uma aplicação na camada de apresentação que simula dados fisiológicos para o módulo integrado.

4.3. Antidote

O Antidote é uma biblioteca *open source* do padrão X73 [Signove 2012]. O principal objetivo da construção dessa biblioteca é manipular a comunicação entre Agent e Manager usando o padrão de comunicação X73. O Antidote disponibiliza um conjunto de APIs que facilita a criação de novas aplicações. A arquitetura do Antidote é ilustrada na Figura 4.

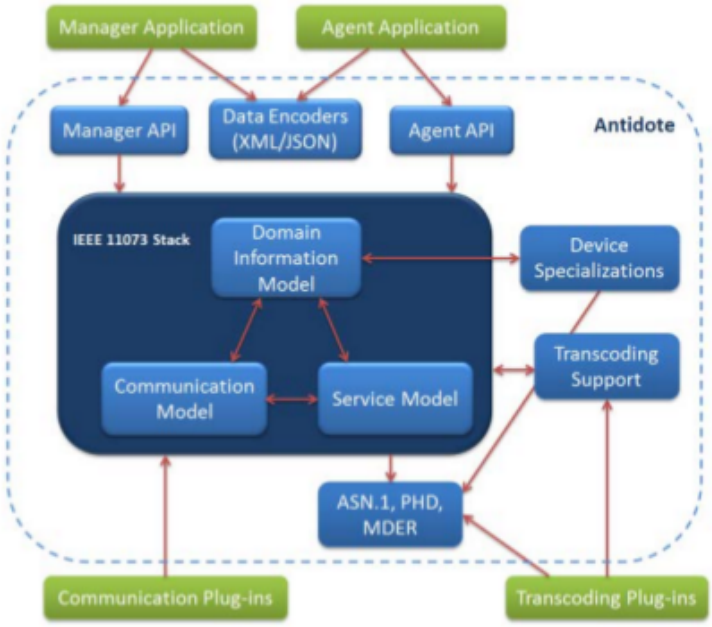


Figura 4. Arquitetura do Antidote. Adaptada de [Signove 2012].

Os componentes em cor verde, *Manager Application*, *Agent Application*, *Communication Plug-ins* e *Transcoding Plug-ins* não fazem parte da biblioteca Antidote. Eles são implementados pelo desenvolvedor da aplicação. Os componentes contidos na área de cor azul fazem parte do padrão X73, incluindo os três principais modelos especificados pelo padrão. Os outros componentes são designados para facilitar o desenvolvimento, entre eles:

- *Manager API e Agent API*, oferecem funções úteis para conduzir uma comunicação usando o padrão X73.
- *Data Encoders*, usados para codificar as medições e configurações em formatos independentes, como XML e JSON. Assim, o desenvolvedor não precisa lidar com os formatos de dados usados no padrão X73.
- *Communication Plug-ins*, oferecem diferentes escolhas para a camada de transporte, permitindo ao desenvolvedor realizar implementações customizadas para cada interface.
- *Transcoding Support*, permite dispositivos que não têm suporte ao padrão X73 se comunicarem com o Antidote.

Para a implementação do módulo adicional ao Plano de Cuidados Ubíquo, foram utilizados exemplos do próprio Antidote, principalmente o *sample_agent_common.c* e o

sample_manager.c. A partir desses que se iniciou a construção da comunicação entre o Agent e o Manager. Outros arquivos foram utilizados, como o *plugin_tcp_agent.c* e o *json_encoder.c* para que a implementação se adequasse à arquitetura do UbiCare.

4.4. Metodologia

A implementação do módulo de captura, manipulação e tratamento dos dados fisiológicos, adicional ao Plano de Cuidados Ubíquo, foi idealizado a partir do estudo dos requisitos de interoperabilidade entre sensores fisiológicos e sistemas ubíquos de assistência domiciliar à saúde. Durante a análise dos requisitos foi constatado que o uso da família de padrões ISO/IEEE 11073 Personal Health Data é recomendável, pois tem suporte por parte da Continua Alliance (veja Seção 3.3) e biblioteca *open-source* para implementação. A partir disso, foi iniciada a investigação e o estudo de aplicação do padrão, considerando uma solução de computação ubíqua.

Para a implementação do módulo, foi usada a arquitetura do UbiCare, juntamente com a solução do Plano de Cuidados Ubíquo, na qual foi estudada e incrementada com uma aplicação de *software* que captura, manipula e trata dois tipos de sensores, pressão arterial e balança.

5. Resultados

A Figura 5 apresenta a aplicação de Plano de Cuidados Ubíquo com o módulo Coletor, proposto e implementado neste trabalho.

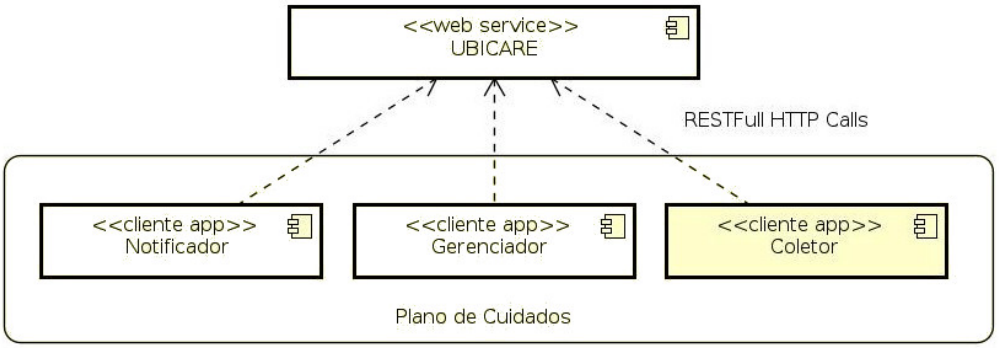


Figura 5. Aplicação de Plano de Cuidados Ubíquo, incluindo o módulo implementado denominado por Coletor.

A aplicação de Plano de Cuidados Ubíquo está estruturada em um conjunto de módulos, representadas na parte inferior da Figura 5. Esses módulos realizam funções necessárias para o Plano de Cuidados Ubíquo:

- Notificador - Notificador do Plano de Cuidados - principal responsável por alertar o paciente do cumprimento de suas prescrições. Essa aplicação foi implementada utilizando dispositivos móveis, TVs e dispositivos Chromecast[3].
- Gerenciador - Gerenciador do Plano de Cuidados – responsável por gerenciar as informações do tratamento do paciente, sendo acessível tanto pelo paciente e pela equipe de saúde, quanto pelos familiares do paciente. Por ser um tipo de aplicação de acesso comum aos usuários do Plano de Cuidados, as informações são apresentadas por ela a partir de um controle de acesso baseado em perfis.

[3] google.com/chromecast

- Coletor - Sensores do Plano de Cuidados - responsável por coletar os dados de sensores fisiológicos, como medidor de pressão arterial e balança. Para isso, o Coletor implementa o padrão de comunicação ISO/IEEE 11073, descrito na Seção 3.

O módulo proposto neste trabalho (Coletor) e integrado ao Plano de Cuidados Ubíquo, recebe os dados de sensores certificados, e os disponibiliza para o gerenciador do Plano de Cuidados (Gerenciador), o qual tem o papel de apresentá-los tanto para o paciente como o profissional de saúde. Além do Coletor, foi desenvolvida uma sessão no módulo gerenciador, que apresenta as informações fisiológicas do paciente. O usuário do sistema pode, então, visualizar os dados de duas formas, tanto em uma tabela ordenada quanto em um gráfico de linhas. O gráfico de linhas proporciona uma análise do comportamento do paciente durante o tratamento. A Figura 6 ilustra a apresentação dos dados fisiológicos.

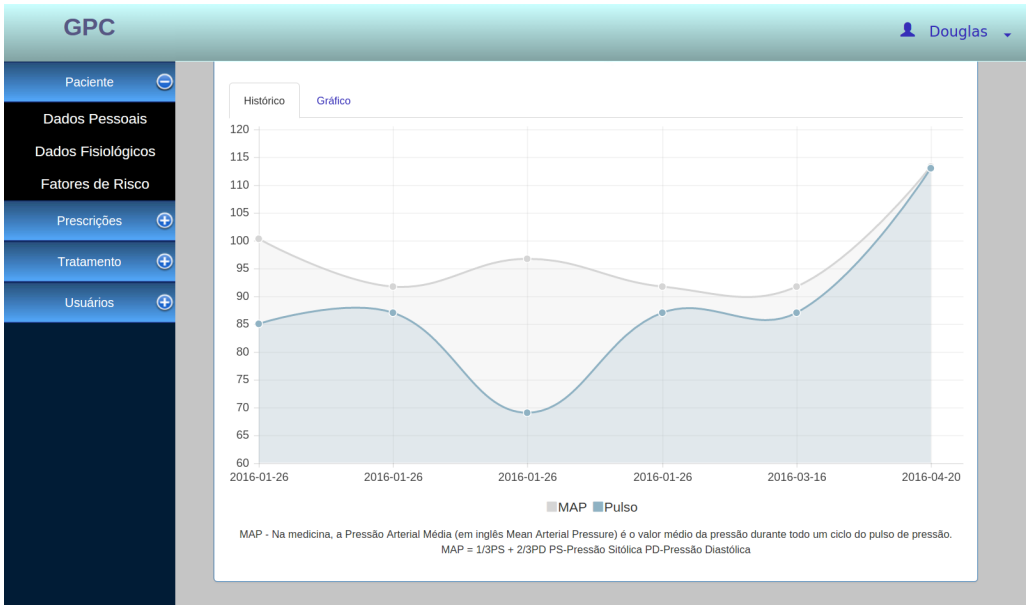


Figura 6. Interface de apresentação dos dados fisiológicos.

Foi desenvolvido também um aplicativo Android que simula um sensor fisiológico, com o objetivo de gerar dados baseados em uma balança e em um medidor de pressão arterial. Os dados gerados são enviados para a aplicação coletora (Coletor), que os encapsula em uma mensagem enviada pelo Agent para o Manager. A Figura 7 apresenta a interface do aplicativo.

6. Conclusão

A solução apresentada neste artigo tem como objetivo geral a aplicação do Padrão ISO/IEEE 11073 Personal Health Data, considerando uma solução de computação ubíqua que utilize sensores fisiológicos em ambiente domiciliar. Isso foi feito por meio da investigação dos requisitos de interoperabilidade em sistemas de assistência domiciliar à saúde, estudo do padrão ISO/IEEE 11073 e implementação de uma solução de *software* que capture, manipule e integre os dois tipos sensores - pressão arterial e balança - ao Plano de Cuidados Ubíquo.

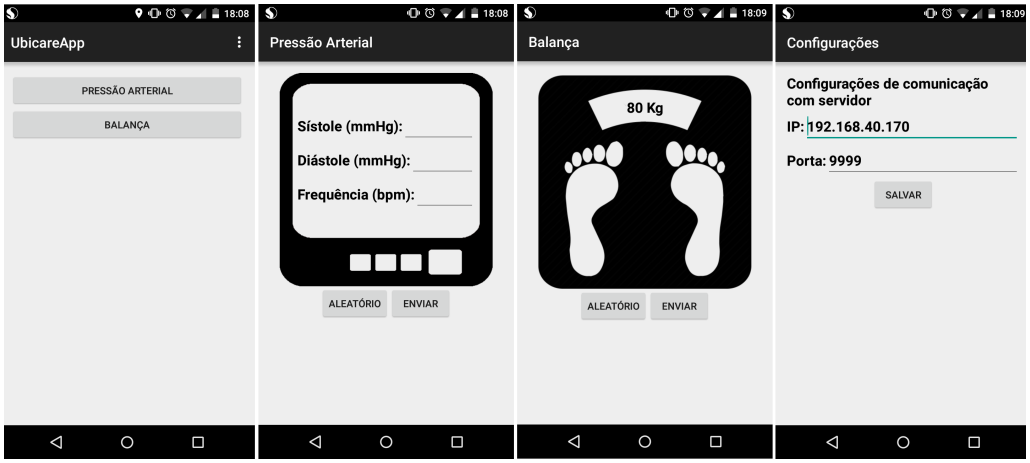


Figura 7. Interface do aplicativo.

Diversos trabalhos têm sido desenvolvidos no sentido de aplicar a tecnologia de computação ubíqua na assistência domiciliar à saúde [Wood et al. 2008, Sztajnberg et al. 2009, He et al. 2013, Kim and Chung 2014, Theoharidou et al. 2014, Brauner et al. 2015, Karthick et al. 2016]. Esses trabalhos, porém, não fornecem soluções de *software* focadas na integração de sensores fisiológicos de uso pessoal, visando a interoperabilidade com outros sistemas de informação em saúde.

Algumas características em relação à proposta apresentada trazem várias facilidades aos pacientes em tratamento, entre elas a facilidade de enviar seus dados fisiológicos ao Plano de Cuidados, acompanhamento de suas informações fisiológicas e a aproximação do profissional de saúde por meio da análise dos dados fisiológicos dos seus pacientes.

Aprimoramentos têm sido feitos ao UbiCare e juntamente ao Plano de Cuidados Ubíquo, no sentido de se seguir os padrões de informática em saúde e as normas de saúde vigentes. Além disso, novas aplicações têm sido desenvolvidas e integradas ao UbiCare, como, por exemplo o uso de redes sociais online como ferramenta de apoio no monitoramento de pacientes [Ribeiro et al. 2016] . Espera-se, assim, o aumento de funcionalidades no contexto de assistência domiciliar à saúde.

Agradecimentos

Os autores agradecem ao CNPQ pelo financiamento parcial deste trabalho.

Referências

Brauner, P., Holzinger, A., and Ziefle, M. (2015). Ubiquitous computing at its best: serious exercise games for older adults in ambient assisted living environments–a technology acceptance perspective. *EAI Endorsed Trans. Serious Games*, 15:1–12.

Copetti, A., Leite, J., Loques, O., Barbosa, T., and da Nóbrega, A. (2008). Monitoramento inteligente e sensıvel ao contexto na assistência domiciliar telemonitorada. *Simpósio Integrado de Software e Hardware, Congresso da SBC*, pages 166–180.

Endrei, M., Ang, J., Arsanjani, A., Chua, S., Comte, P., Krogdahl, P., Luo, M., and Newling, T. (2004). *Patterns: service-oriented architecture and web services*. IBM Corporation, International Technical Support Organization.

Germano, E., Battisti, D., Ribeiro, H., and Carvalho, S. (2016). Plano de cuidados ubíquo para acompanhamento domiciliar de pacientes. Aceito para publicação (to appear). Congresso Brasileiro de Informática em Saúde (CBIS).

Germano, E., Carvalho, S., and De Souza-Zinader, J. (2015). Plano de cuidados ubíquo com sistema de notificações voltado a pacientes domiciliares. *III Escola Regional de Informática de Goiás*, pages 33–44.

Haynes, R. B., McDonald, H. P., and Garg, A. X. (2002). Helping patients follow prescribed treatment: clinical applications. *JAMA*, 288(22):2880–2883.

He, C., Fan, X., and Li, Y. (2013). Toward ubiquitous healthcare services with a novel efficient cloud platform. *IEEE Transactions on Biomedical Engineering*, 60(1):230–234.

Karthick, K., Praveen, T., and Vidya, M. Q. M. (2016). Enabling human health care monitoring using ubiquitous computing. *Indian Journal of Science and Technology*, 9(16).

Kim, J. and Chung, K.-Y. (2014). Ontology-based healthcare context information model to implement ubiquitous environment. *Multimedia Tools and Applications*, 71(2):873–888.

Martinez, I., Fernandez, J., Galarraga, M., Serrano, L., de Toledo, P., Jimenez-Fernandez, S., Led, S., Martinez-Espronceda, M., and Garcia, J. (2007). Implementation of an end-to-end standards-based patient monitoring solution. *IEE Proceedings Communications - Special Issue on Telemedicine and e-Health Communication Systems COM*.

Piniewski, B., Muskens, J., Estevez, L., Carroll, R., and Cnossen, R. (2010). Empowering healthcare patients with smart technology. *Computer*, 43(7):27–34.

Ribeiro, H., Battisti, D., Germano, E., and Carvalho, S. (2016). Notificações de monitoramento remoto de pacientes usando redes sociais. Aceito para publicação (to appear). Congresso Brasileiro de Informática em Saúde (CBIS).

Signove (2012). *Antidote: Program Guide. Documentation for developers of applications based on Antidote IEEE 11073 library*. Signove, 2.12 edition. http://oss.signove.com/index.php/File:AntidoteProgramGuide.pdf, acessado no dia 12/07/2016.

Sztajnberg, A., Rodrigues, A. L. B., Bezerra, L. N., Loques, O. G., Copetti, A., and Carvalho, S. (2009). Applying context-aware techniques to design remote assisted living applications. *International Journal of Functional Informatics and Personalised Medicine*, 2(4):358–378.

Theoharidou, M., Tsalis, N., and Gritzalis, D. (2014). Smart home solutions for healthcare: privacy in ubiquitous computing infrastructures. *Handbook of smart homes, health care and well-being*.

Weiser, M. (1991). The computer for the 21st century. *Scientific american*, 265(3):94–104.

Wood, A. D., Stankovic, J. A., Virone, G., Selavo, L., He, Z., Cao, Q., Doan, T., Wu, Y., Fang, L., and Stoleru, R. (2008). Context-aware wireless sensor networks for assisted living and residential monitoring. *IEEE network*, 22(4):26–33.

Yao, J. and Warren, S. (2005). Applying the iso/ieee 11073 standards to wearable home health monitoring systems. *Journal of Clinical Monitoring and Computing*, 19(6):427–436.

Yu, L., Lei, Y., Kacker, R. N., Kuhn, D. R., Sriram, R. D., and Brady, K. (2013). A general conformance testing framework for ieee 11073 phd's communication model. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, page 12. ACM.

Projeto Redação: Criação de Valor por meio da Aplicação do Modelo de Comunidade de Prática

Marcelo Akira Inuzuka[1], Valéria Maria Rosa[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

marceloakira@inf.ufg.br, valeria.si.ufg@gmail.com

Resumo. *Este trabalho teve como objetivo a análise do Projeto Redação, um sistema colaborativo online de preparação para a prova de redação do Exame Nacional do Ensino Médio (ENEM)" quanto Comunidade de Prática (CoP), que ocorre quando duas ou mais pessoas se unem para aprender ou tentar aprender algo. Houve também a aplicação de suas principais características para criação de valor. Realizou-se uma pesquisa entre os membros do Projeto, com o intuito de aumentar a criação de valor, e por meio disso, foram realizadas algumas sugestões de implementação para aumentar o engajamento e interação entre os membros. Essas sugestões foram avaliadas pela área estratégica do Projeto Redação. O feedback recebido foi positivo, e a empresa adicionará as sugestões de implementação no próximo sprint.*

Abstract. *This work had as objective the analysis of Projeto Redação, a which occurs when two or more people come together to learn or try to learn something and its main features to create value. So, it was verified that the project answered the characteristics of a CoP, and through its degree of maturity, a survey among the members of the project, in order to increase value creation. A survey among the members of the project, in order to increase engagement and interaction between the members. These suggestions were evaluated by strategic area of Projeto Redação. The feedback received was positive, and the company will put the suggestions for implementation to be carried out in the next sprint.*

1. Introdução

A plataforma web de correção de redações no padrão ENEM, Projeto Redação[1], analisada neste trabalho foi criada após a grande mudança ocorrida na educação brasileira no que se refere ao ingresso em instituições de ensino superior. Antes, cada instituição de ensino superior tinha seu próprio método de ingresso, sendo necessário que o aluno prestasse vestibular separadamente para cada instituição desejada. Com a mudança, a maioria das instituições passaram a adotar uma única prova para seleção - o Exame Nacional do Ensino Médio (ENEM).

Uma pesquisa realizada pela FUVEST[2] baseada nas respostas do questionário socioeconômico dos vestibulandos, mostra que, cada vez mais, os alunos estão deixando de fazer cursos pré-vestibulares. Segundo essa pesquisa, o índice de alunos que não fizeram curso preparatório subiu de 29,4% em 2009, para 37,5% em 2013. Isso levou muitos alunos recorrerem a cursos preparatórios online, como o Projeto Redação.

A Comunidade de Prática (CoP), expressão criada por Jean Lave e Etienne Wenger, em 1987, é um grupo de pessoas que possuem um interesse ou problema em comum. Essas pessoas se encontram para que possam agregar valor por meio de interações: elas compartilham informações, e conselhos. Assim, elas acumulam conhecimento agregando valor no aprendizado que encontraram juntas. Elas desenvolvem relações pessoais e instituem formas de interação.

Analisamos as características de comunidade de prática no Projeto Redação (PR), assim como a criação de valor por meio dos ciclos, enfatizando o impacto da aplicação do modelo de Comunidade de Prática na compreensão do processo de criação de valor no Projeto Redação.

Para a análise desse impacto foi aplicado um questionário com algumas questões chave de criação de valor. Responderam este questionário 53 usuários ativos no PR e que já haviam ou enviado uma redação ou um comentário. Foi feita a análise dessas respostas para assim ter sobre a criação de valor. Assim, foram feitas algumas sugestões para melhorar o desempenho da comunidade, como ranking/pódio, a utilização de moderadores e de fóruns. Essas sugestões foram passadas para a parte estratégica do PR, que avaliou a importância da implementação dessas características.

2. Modelagem

2.1. Comunidade de Prática

A construção do conceito de CoP[3] é dada por um conceito central, a teoria social da aprendizagem, que agrega mais quatro referencais teóricos: teorias da identidade, da prática, da experiência situada e da estrutura social. Possui também quatro linhas teóricas: teoria do significado, da subjetividade, do poder, e da coletividade.

A CoP apresenta três características principais: domínio, comunidade e prática. O domínio é o que cria a base comum e o senso de identidade. A comunidade se refere ao engajamento mútuo, participação de discussões e trocas de informação. A prática é o desafio que une e mobiliza os membros, criando uma prática compartilhada.

A maturidade realiza a conexão entre atividades de uma CoP e à melhoria de desempenho das organizações por meio de uma avaliação de valor ao longo de cinco ciclos: de valor imediato (como conversas online e fornecimento de dicas), potencial (que pode assumir a forma de capital humano, social, tangível, de reputação e de aprendizagem), aplicado (adaptação do capital e aplicação em uma situação específica), realizado (quando novas práticas ou ferramentas aplicadas não são suficientes) e de resignificação (reformulação de estratégias, objetivos e valores).

A criação de valor acontece por meio dos cinco ciclos, e para isso, são coletadas histórias de criação de valor. Elas permitem uma imagem mais confiável de como é realizada a criação de valor, e são contadas pelas pessoas envolvidas na comunidade.

2.2. Sobre o Projeto Redação

O Projeto Redação é uma plataforma de correção de redação nos padrões do ENEM. Ele tem como objetivo ajudar alunos a ter um bom desempenho na redação do ENEM, que equivale a 1/5 da nota total desse exame. A platafotma além de fornecer a interação

aluno-professor, conta com a interação aluno-aluno, promovendo a aprendizagem colaborativa. Na interação aluno-professor, a redação será corrigida por um profissional que orientará o aluno identificando seus pontos fortes e fracos. No segundo caso, o aluno aprende com a redação e correção de outros alunos da plataforma, além de poder opinar em outras redações. De acordo com a própria descrição do site, "O Projeto Redação é um site de correção profissional e colaborativa de redações. Você estuda praticando sua escrita e lendo redações de outros estudantes."

No modo colaborativo, quando uma redação é enviada por um aluno ela estará automaticamente disponível para todos os usuários. Já no modo profissional, a redação só sera disponibilizada publicamente caso o autor permita. Caso o aluno não queira ser identificado, a redação poderá ser enviada de modo anônimo.

Após a redação ser disponibilizada publicamente no site, outros usuários da plataforma podem fazer comentários sobre o texto. Assim, além de ajudar o autor da redação, o usuário que comenta também está exercitando seu conhecimento.

2.3. Aplicação da CoP

No Projeto Redação há a necessidade de retenção e troca de conhecimento entre os colaboradores com conhecimentos e experiências distintas, no que se refere aos alunos que buscam alcançar uma boa nota da redação do ENEM. Com base na análise dos textos enviados pelos alunos à plataforma, percebe-se que eles possuem conhecimentos em diferentes graus (alguns textos são bem estruturados e com poucos erros, enquanto outros possuem muitos problemas com coesão e coerência) e pelas notas recebidas na correção profissional (o professor avalia o texto de acordo com as cinco competências do ENEM, atribuindo uma nota de 0 a 200 para cada uma delas). E cada um apresenta uma experiência distinta, no que se refere a toda vida estudantil - há alunos que estão prestando o exame pela quinta vez, enquanto outros prestarão pela primeira vez. Além disso, é preciso criar vínculos emocionais entre os alunos para que eles sempre busquem ajudar os outros, um dos pilares definidos por esse tibo de abordagem (CoP).

Assim, temos um mesmo desafio compartilhado por todos os membros do PR, um dos elementos centrais da CoP: conseguir uma boa nota na redação, para que assim entrem em uma universidade pública. O site já apresenta uma área onde os alunos possam colaborar com os outros na aprendizagem: uma área de comentários nas redações.

Para ser considerada uma comunidade de prática, a comunidade deve apresentar os seguintes elementos (Wenger, 2009), que são contemplados também pelo PR:

- Domínio: O interesse em comum é o ingresso em uma faculdade pública por meio de uma boa nota na redação. Eles possuem um compromisso com a comunidade até conseguirem atingir esse objetivo.
- Comunidade: os alunos participam de discussões em conjunto, por meio dos comentários feitos na redação.
- Prática: Os membros da comunidade são alunos que estão no ensino médio, cursos pré-vestibulares ou que estudam em casa. Eles já possuem experiências com produção de texto, e alguns com a prova do ENEM. Assim, eles trocam experiências e conhecimento.

Para uma avaliação mais criteriosa sobre a aplicabilidade do modelo de CoP (Comunidade de Prática), foram aplicadas entrevistas entre os membros da comunidade.

De acordo com os estágios de desenvolvimento de uma comunidade de prática, o Projeto Redação classifica-se como estando no primeiro estágio, pois reune pessoas com questões e necessidades similares: alunos com dificuldade em fazer uma boa produção textual e com necessidade em passar em uma universidade pública.

Quanto aos graus de participação, pode-se de dizer que o PR possui um baixo número de comentários quando comparado ao número de usuários ativos. São 28.085 comentários e 56.556 usuários ativos, dando uma média de 0.4966 comentários/usuário (supond que casa usuário tenha enviado um comentário), ou seja, menos de 50% dos usuários colaboraram. No entanto, não se pode afirmar que existe grau de participação periférico para todos usuários. Seria necessáro criar e identificar percentuais de classes de graus de participação, baseado em critérios próprios, para ajudar a analisar a tomada de decisão futura.

Essa análise, serve à diretoria do PR visto que, esta deve analisar ou julgar se há baixo ou alto nível de engajamento e se há necessidade de criar ações para aumento de engajamento.

2.4. Criação de Valor no Projeto Redação

Para investigar formas de criar valor no PR, algumas questões podem ser abordadas de acordo com cada ciclo. Foi enviado *e-mail* contendo essas questões para 200 membros ativos, tendo-se obtido 53 questionários respondidos. Essa escolha foi realizada pois o questionário aborda a experiência do membro na comunidade, assim como suas interações.

Nesse questionário, que segue abaixo, foram utilizadas algumas questões chave, retiradas do *Promoting and assessing value creation in communities and networks: a conceptual framework*[4] para criação de valor.

1. Foi divertido, inspirador?
2. Quão relevante foi a interação de comentários para você?
3. Ao colaborar, você adquiriu novas habilidades e conhecimentos?
4. Seu entendimento do assunto ou sua perspectiva mudaram?
5. Ganhou acesso a novas pessoas?
6. Pode confiar o suficiente em outras pessoas para recorrer a elas em busca de ajuda?
7. Se sente menos isolado?
8. Teve acesso a informações que não teria de outra maneira?
9. Você vê oportunidades para aprender que não via antes?
10. Onde aplicou as habilidades que adquiriu?
11. Quando e como usou um documento ou ferramenta que o Projeto Redação produziu ou tornou acessível?

3. Resultados

As estatísticas das respostas de cada pergunta do questionário aplicado entre os membros ativos do Projeto Redação foram:

1. 88.2% dos membros consideram divertido/inspirador a atividade de comentários no PR.
2. 83% acharam relevante a interação. Abaixo, temos algumas respostas dos membros:

   "Muito relevante! Já que em todas as redações que comentei, os seus respectivos donos também comentaram nas minhas redações. Assim trocamos informações."

   "Extremamente baixo, percebo que dificilmente alguém comenta em alguma redação. O que vocês deveriam fazer era colocar as barras de opções para as pessoas darem as notas por competência, porque fazer um comentário exige mais esforço e disposição do indivíduo que está lendo. Se o indivíduo pudesse só dar sua nota para as redações que ele lesse seria muito mais interativo, e também seria muito melhor para a pessoa que enviou a redação"

3. 88.6% consideraram que adquiriram novas habilidades e conhecimentos ao comentarem. Abaixo, segue a resposta de um membro sobre essa questão:

   "Sim, para ter certeza de que meu comentário não levantaria uma observação errada ou inadequada tive que pesquisar, estudar antes de elaborar meu comentário."

4. 86.8% dos membros consideram que seu entendimento do assunto ou perscpectiva mudaram.
5. 46.3% alegaram que, por meio do PR, ganharam acesso a novas pessoas.
6. 75.5% confiam o suficiente em outras pessoas para recorrer em busca de ajuda.
7. 77.8% dos membros se sentem menos isolado.
8. 72.2% tiveram acesso a informações que não teriam de outra maneira.
9. 98.1% viram oportunidades para aprender que não viam antes;
10. Sobre onde os membros aplicaram as habilidades adquiridas, segue abaixo algumas respostas:

    "Nas provas de vestibulares que prestei."

    "Nos processos intrapessoais do cotidiano."

    "Na minha forma de escrever e, também, no convívio com meu grupo de estudos."

11. Quanto a quando e como usou um documentos ou ferramenta do PR:

> "Faço uso de todo material que chega até mim através de email do "grupo" do Projeto Redação"
>
> "Sempre, estou aprendendo a fazer redações com o projeto, e minhas notas estão melhorando :D"
>
> "Eu uso os textos motivadores para testar técnicas e me familiarizar como exame do Enem."

Percebe-se que muitos membros sentem-se engajados no projeto, porém muitos ainda reclamam de que não tiverem um feedback nas suas redações, fato comprovado pela quantidade de comentários e redações enviadas: 28.085 e 38.665, respectivamente. Assim, precisa-se encontrar meios de fazer com que essa interação de comentários aumente.

Para aumentar o engajamento dos membros e aumentar a criação de valor, primeiro recomenda-se incentivá-los a comentar mais. Uma das ideias que poderia ser aplicada é a criação de um pódio na plataforma, informando quais membros mais colaboraram com comentários relevantes, e até mesmo, oferecendo premiações para os primeiros colocados. A comunidade "Viva o Linux", criada em 2002, já faz uso desse recurso, e possui 200.429 membros, 239.124 comentários e 5.983 artigos.

Outro conceito importante de CoP empregado na comunidade "Viva o Linux", e que pode ser utilizado no PR é a existência de moderadores: pessoas com maior experiência e conhecimento em determinada área e que podem orientar outros membros. No caso, pessoas com maior quantidade de comentários relevantes nas redações podem ser convidados a serem moderadores do projeto, e assim, colaborar de forma mais significativa com os outros membros.

Outra característica aconselhada a se aplicar no projeto é o fórum. Nele, o membro pode enviar perguntas e até mesmo colocar informações que considera relevantes para outros membros. A colaboração do membro pode ser medida também por meio de contribuições no fórum, e não somente em comentários na redação.

Sugeriu-se para a diretoria do Projeto Redação a implementação das áreas mencionadas anteriormente: Ranking/Pódio, eleição de membros moderadores na comunidade e fórum. A proposta foi analisada e teve o seguinte *feedback*:

"Sempre acreditei que a saída para que conseguíssemos expandir o uso de forma viral[1] da comunidade do Projeto Redação seria por meio de gamificação[2]. Porém, para implantar o desenvolvimento de funcionalidades que suportassem essa ideia, teríamos que alocar recursos que no nosso caso, já são limitados. Decidimos começar as operações do Projeto Redação como um negócio sustentável financeiramente desde o começo.

E como correção de redação não é um negócio altamente lucrativo e escalável, resolvemos focar em melhorias do sistema para que os corretores pudessem corrigir de

[1] Termo usual da internet que designa a ação de fazer com que algo se espalhe rapidamente.

[2] Aplicar técnicas de engajamento dos jogos nas mais diferentes situações.

maneira mais rápida e eficiente, ou seja, tivemos que sacrificar, por hora, o desenvolvimento da comunidade.

As propostas sugeridas pelo trabalho atendem perfeitamente nosso interesse, e vamos colocar em nossa lista de tarefas para os próximos ciclos de sprint[3].

Atrelado ao desenvolvimento, também vamos focar nas métricas de utilização (engajamento), que vão sustentar a hipótese de que o desenvolvimento das melhorias vai aumentar a interação entre os membros e, consequentemente, o aprendizado daqueles que utilizam o site como fonte de estudo e conhecimento.

Do ponto de vista de negócio, aumentar o engajamento dos alunos para utilizar nossa ferramenta de comunidade também significa possibilidade aumentar nossa taxa de usuários gratuitos para usuários pagantes, considerando que temos um modelo freemium. Com o freemium[4], uma grande base de usuários do serviço gratuito é subsidiada pela pequena base de usuários premium e por anunciantes."

Pelo relato do diretor estratégico, percebemos que há valor para a organização a implementação das características de CoP, pois isso traria tanto maiores benefícios para os usuários do PR quanto para o negócio em si.

4. Conclusão

Este trabalho realizou a análise do Projeto Redação, uma plataforma web para correção de redações no padrão ENEM, quanto Comunidade de Prática para a criação de valor. Assim, primeiro verificou-se se o PR contemplava as três principais características que definem uma CoP: domínio, comunidade e prática. A verificação foi feita de acordo com a análise de dados oriundos do próprio PR.

Para investigar as formas de criação de valor, realizou-se uma pesquisa entre os membros. A pesquisa estava toda voltada para a área de comentários nas redações. Percebeu-se um bom índice em todas as perguntas, sendo que apenas uma delas, a de obter acesso a novas pessoas, obteve índice menor que 50%. Embora com bons resultados, entre os pontos fracos mais frequentes apontados pelos usuários, destacou-se o de baixo número de comentários: muitas redações não haviam recebido feedback de outros usuários.

Para resolver os problemas encontrados, foram sugeridas algumas medidas, as quais já possuem aplicação em comunidades bem sucedidas, como o "Viva o Linux"[5]. A implentação de um ranking/pódio que informa os membros que mais colaboraram com comentários relevantes, incentivaria os membros a comentarem mais, para que assim tivessem reconhecimento na plataforma - a premiação aos primeiros colocados desse ranking, atrairia mais ainda. Outra ideia sugerida foi a de selecionar alguns membros para serem moderadores. Quando um membro tivesse alguma dúvida, ele já teria a quem recorrer sem precisar esperar alguém comentar em sua redação. Além

[3] Time Box dentro do qual um conjunto de atividades deve ser executado.

[4] Termo usual da internet que designa a ação de fazer com que algo se espalhe rapidamente.

disso, sugeriu-se a criação de fóruns para postagens de assuntos que os participantes consideram importantes e para questionamentos

O trabalho de análise foi passado para a área estratégica da empresa: a avaliação foi considerada positiva, onde reconheceu-se a importância de maior engajamento entre os membros, tanto para o maior aprendizado, quanto para maior conversão de usuários gratuitos para pagantes. Reconheceu-se assim o impacto na aplicação do modelo de CoP no processo de criação de valor no Projeto Redação.

5. Referências

[1] Projeto Redação, 2016. Disponível em: <https://www.projetoredacao.com. br/>. Acesso em: 12 de julho de 2016.

[2] TOLEDO, L. F. 4 em cada 10 passam na fuvest sem cursinho, nov 2014.

[3] CABELLEIRA, D. M. Comunidades de prática – conceitos e reflexões para uma estratégia de gestão do conhecimento. XXXI EnANPAD, 31:101–116, 2007

[4] ETIENNE WENGER, BEVERLY TRAYNER, M. D. L. Promoting and assessing value creation in communities and networks: a conceptual framework. The Netherlands, 2011.

[5] Viva o linux, 2016. Disponível em: <https://www.vivaolinux.com.br/>. Acesso em: 28 de junho de 2016.

Dados abertos da cultura: uma análise sob a ótica da tecnologia da informação

Lafaiet Castro e Silva[1], **Núbia Rosa da Silva**[2], **Douglas Farias Cordeiro**[1]

[1]Faculdade de Informação e Comunicação - Universidade Federal de Goiás
Campus Samambaia – 74.590-900 – Goiânia – GO – Brasil

[2]Instituto de Biotecnologia - Universidade Federal de Goiás
Av. Dr. Lamartine P. de Avelar, 1200 – 75704-020 – Catalão - GO – Brasil

lafaietsilva@gmail.com , {nubia,cordeiro}@ufg.br

Abstract. *In an information society, where the need to explore data for the generation of knowledge and decision-making, coupled with the growing demand for transparency in what refers to public policy, opening government data becomes an important and interesting alternative by providing the reuse of data in a variety of purposes. Nevertheless, the opening data is still considered a major challenge, both as refers to the management aspects and regarding the technological factors. In this article, an analysis of the open government data from the perspective of information technology is presented, discussing results of a development strategy and implementation of an API for opening data of a cultural project management system of the Ministry of Culture of Brazil.*

Resumo. *Em uma sociedade da informação, onde a necessidade de explorar dados para a geração de conhecimento e tomada de decisão, aliada à crescente busca pela transparência no que refere-se às políticas públicas, a abertura de dados governamentais torna-se uma importante e interessante alternativa, por proporcionar o reuso dos dados sob os mais diferentes propósitos. Apesar disso, a abertura de dados ainda é considerada um grande desafio, tanto no que refere-se à aspectos de gestão, quanto à fatores tecnológicos. Neste artigo será apresentada uma análise sobre a abertura de dados governamentais sob o olhar da tecnologia da informação, onde serão discutidos resultados de uma estratégia de desenvolvimento e implantação de uma API para abertura de dados de um sistema de gestão de projetos culturais do Ministério da Cultura do Brasil.*

1. Introdução

De forma geral, um dado pode ser considerado qualquer elemento que se refira à representação de algo, sendo originalmente apresentado em um formato "bruto", aparentemente sem valor, mas que a partir de intervenções específicas pode ser transformado em informação, a qual, por sua vez, gera conhecimento e valor, podendo ser considerada, em diversos casos, como um dos principais ativos de uma organização [Choo 2003]. Estes conceitos estão intrinsecamente ligados à tecnologia, a qual, através de avanços alcançados pela Ciência da Computação e áreas correlatas, permitiram um aumento exponencial na manipulação, recuperação e utilização de dados, sob as mais diferentes perspectivas, fato este que aumentou ainda mais o interesse naquilo que se refere à dados e informação.

Embora no âmbito de várias organizações existam tipos específicos de dados que necessitam ser resguardados de divulgação e acesso não autorizado, existem também aqueles passíveis de publicação e reutilização para diferentes fins, os quais são conhecidos como dados abertos. Este tipo de dado ganha uma especial atenção quando coletados, mantidos e utilizados pelo Governo, tanto pelo aspecto da centralidade, quanto também pelo direito de acesso, garantido constitucionalmente. De acordo com [Open Knowledge Foundation 2011]:

"Dado aberto é aquele que pode ser livremente utilizado, reutilizado e redistribuído por qualquer pessoa, e deve, no máximo, atribuir à fonte original e/ou fazer o compartilhamento desses dados utilizando a mesma licença em que as informações foram apresentadas".

De forma geral, a discussão sobre dados abertos pode ser focada, de acordo com o disposto por [W3C 2011], através dos seguintes pontos específicos:

- **Disponibilidade de acesso** – o dado deve cumprir ao requisito de disponibilidade integral, através de um custo razoável de reprodução, ou preferencialmente, sem custo, estando acessível por meio de download na Internet, e sendo apresentado em formato que seja conveniente e modificável;
- **Reutilização e redistribuição** – a reutilização e redistribuição deve ser garantida, assim como o cruzamento com outros conjuntos de dados.
- **Participação universal** – qualquer indivíduo deve ter o direito de uso, reuso e redistribuição, sendo vedado qualquer tipo de discriminação, sejam contra áreas de atuação, pessoas ou grupos específicos.

Os dados abertos, para serem classificados como tal, também deve obedecer a oito princípios básicos. Estes princípios foram definidos por um conjunto de estudiosos da área, formado por colaboradores provenientes de universidades, do governo e de empresas da área de tecnologia, com o objetivo principal de debater e desenvolver um entendimento colaborativo a cerca dos conceitos envolvidos à Governo Aberto, e as ligações deste com a democracia. Neste contexto, para que dados sejam então descritos como abertos, os mesmos devem ser:

- **Acessíveis** – disponibilizados para o maior número possível de pessoas, atendendo, assim, aos mais diferentes propósitos possíveis;
- **Atuais** – é necessário que os dados, para que preservem seu valor, sejam publicados como o menor intervalo possível após a sua criação, fato este que encontra-se embasado nos termos de que, quanto mais recentes e atuais são os dados, mais úteis poderão ser a seus usuários;
- **Amparados por licenças livres** – os dados não podem estar submetidos a direitos autorais, patentes, marcas registradas ou regulações de segredo industrial;
- **Completos** – qualquer dado público deve ser disponibilizado, sendo que, obedecendo a princípios de privacidade e segurança da informação, um dado público pode ser descrito como aquele que não está sujeito a restrições de privacidade, segurança ou outros privilégios;
- **Compreensíveis por máquina** – os dados devem ser apresentados em uma estrutura que possibilite seu processamento automático. Por exemplo, um conjunto de dados apresentado no formato de uma tabela em um arquivo PDF pode ser

facilmente compreendido por pessoas, entretanto, a sua compreensão, em termos de manipulação e processamento, por uma máquina torna-se algo mais complexo. Por outro lado, um conjunto de dados dispostos em uma tabela em formato estruturado, como CSV, pode ser processado de maneira direta e precisa por sistemas computacionais orientados à este propósito;

- **Não discriminatórios** – qualquer indivíduo pode ter acesso aos dados, sem que exista necessidade de cadastro ou de qualquer procedimento que possa impedir o acesso;
- **Não proprietários** – os dados não podem ser controlados, exclusivamente, por nenhuma entidade ou organização;
- **Primários** – os dados devem ser apresentados exatamente da forma com que foram colhidos da fonte, com o maior nível possível de granularidade, sem agregação ou modificação. Neste contexto, é importante destacar que um dado, mesmo após um processo de manipulação ou processamento, não garante a geração de informação, podendo gerar um outro tipo de dado.

2. Dados Abertos Governamentais

De acordo com [Agune et al. 2010], um governo aberto pode ser descrito como aquele que pauta entre suas políticas pela disponibilização, por meio da rede mundial de computadores, aqueles dados que sejam passíveis de classificação como dados de domínio público, permitindo, a partir disso, conhecimento e uso pela sociedade. Neste contexto, a OGP - *Open Government Partnership* (do inglês, Parceria para Governo Aberto) descreve que um governo aberto deve alcançar os seguintes requisitos:

- Proporcionar um aumento na disponibilidade de informações a cerca de atividades de governo;
- Fornecer apoio à participação social;
- Proporcionar a implementação dos mais elevados padrões de integridade profissional no campo da Administração;
- Promover o acesso a tecnologias que busquem a transparência e a prestação de contas.

No âmbito nacional, constitucionalmente, todo o cidadão brasileiro possui direito garantido de acesso às informações públicas, em qualquer uma das três esferas de poder. Este direto é relatado no texto da Lei 12.527, de 18 de novembro de 2011, conhecida como Lei de Acesso à Informação. Esta lei está diretamente ligada às políticas de abertura de dados por parte dos orgãos governamentais. Neste contexto, a Figura 2 apresenta aspectos relevantes ligados à abertura de dados governamentais.

2.1. Licenças de Utilização

Questões relacionadas ao licenciamento de dados abertos como forma de garantia de uso e reuso livre sempre estiveram pautadas como um dos pontos centrais de discussão em movimentos de dados abertos, fato este justificado por uma das próprias características dos dados abertos: a possibilidade de explorá-los de forma livre sem necessidade de solicitações ou identificação. Neste contexto, embora exista uma tênue ligação entre dados e sistemas, é importante destacar que licenças de software livre ou de conteúdo livre não se encaixam no termos de dados abertos, uma vez que estes tratam-se de ativos de natureza

Figura 1. Potencial dos dados abertos.

distinta. A partir disso, um conjunto de licenças específico foi concebido com o propósito de tratar aspectos relacionados a dados e bancos de dados, o qual é conhecido como *Open Data Common*.

O conjunto de licenças *Open Data Common* segue a seguinte estrutura:

- **Open Database License** (ODbL) – licença voltada à arquitetura da informação, esquemas de bancos de dados e formas de organização dos dados, exigindo atribuição de autoria e compartilhamento;
- **Database Content License** (DbCL) – possui as mesmas exigências apresentadas pela ODbL, porém é aplicável ao conteúdo dos bancos de dados, ou seja, os próprios dados em si;
- **Open Data Commons Attribution License** – licença aplicável aos dados e a bancos de dados, apresentando unicamente a exigência de atribuição de autoria;
- **Public Domain Dedication and License** (PDDL) – licença aplicável aos dados e aos bancos de dados, porém com renúncia de todos os direitos, ou seja, os ativos licenciados por esta são considerados de domínio público.

É fundamental destacar que, embora as licenças apresentadas anteriormente consigam cobrir grande parte das necessidades relacionadas à dados e bases de dados, não existem licenças específicas para dados abertos.

3. Gerenciando a abertura de dados

O processo de abertura de dados deve estar focado essencialmente na simplicidade. Inicialmente, o foco é o desenvolvimento de um plano pequeno, simples e rápido. Neste sentido, é importante notar que não existe uma obrigatoriedade de abertura de todo o conjunto de dados de uma organização de forma única, sendo que, de acordo com o [W3C 2011], uma das melhores estratégias é selecionar somente um conjunto de dados para abertura,

ou mesmo um subconjunto de um grande conjunto de dados. Essa abordagem proporciona um processo mais gerenciável, facilitando as fases de testes e receptividade por parte da sociedade. Além disso é importante destacar que trabalho ágil pode ser crucial nos resultados obtidos, uma vez que, por se tratar de um processo de inovação, isso poderá auxiliar na detecção de falhas e sucessos.

Outro ponto de fundamental atenção refere-se ao tratamento de riscos e ameaças relacionados à cultura organizacional da instituição, principalmente no refere-se à abertura de dados governamentais, onde, normalmente, tem-se aspectos extremamente particulares de cultura organizacional, tais como concentração de poder, burocracia, descontinuidade, entre outros [de Souza Pires and Macêdo 2006]. Além disso, durante este processo, é fundamental que mesmo as partes que não estão diretamente ligadas à abertura de dados, tais como as comunidades externas à organização, devem ser envolvidas e consideradas, fato este que deve já ser levado em conta deste o início de um projeto de abertura de dados.

Ainda neste contexto, de acordo *Open Data for Development in Latin America and the Caribbean*, o ciclo de abertura de dados, seja no governo ou não, é amplo e precisa ser observado para que as iniciativas de publicação e abertura de dados sejam sustentáveis. Em outras palavras, para que os benefícios sinalizados pelos estudiosos que propagam os conceitos dos dados abertos se tornem algo palpável para a humanidade, é necessário observar todo o processo que envolve a publicação de dados abertos na Web.

4. Publicação de Dados

Um dos principais aspectos inerentes às propriedades dos dados abertos é a sua possibilidade de reutilização. Neste sentido, para que tal fator seja satisfeito, é necessário que os dados estejam estruturados e disponibilizados em um formato que permita o seu reuso. Entre as principais alternativas, podem ser destacados os formados: JSON, XML, CSV e RDF, os quais podem ser gerados através da utilização de um API desenvolvida para tal propósito. A seguir são apresentados os principais aspectos sobre estes e outros formatos.

CSV

O formato CSV (*Comma-Separated Values*), que literalmente significa "Valores separados por vírgulas", se refere a um formato de arquivo suportado por praticamente todos os softwares de planilhas eletrônicas, assim como por sistemas gerenciadores de bancos de dados (SGBD). Este formato é consideravelmente popular, sendo reconhecido internacionalmente.

No Brasil, o padrão CSV foi adotado pelo e-Ping[1] para tratamento de arquivos do tipo "banco de dados", sendo definido para utilização no Portal da Transparência (`http://www.portaldatransparencia.gov.br/`), em conformidade com a Lei de Acesso à Informação (Lei nº 12.527/2011).

TSV

[1]Padrões de Interoperabilidade de Governo Eletrônico que define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação (TIC) no governo federal, estabelecendo as condições de interação com os demais poderes e esferas de governo e com a sociedade em geral.

O formato TSV (*Tab-Separeted Values*), ou valores separados por tabulação (TSV), é um formato ligeiramente semelhante ao CSV, podendo também ser criado ou visualizado pela maior parte dos softwares de planilha e editores de texto. As principais regras do TSV são:

- Cada entrada deve ocupar uma única linha;
- A primeira linha se refere à linha de cabeçalho, que rotula cada campo;
- Um campo deve conter dados, como um número ou texto.
- Os campos são separados por tabulações.
- Cada linha deve conter exatamente o mesmo número de campos.

XML

De acordo com o [W3C 2011], XML (*Extensible Markup Language*) ou, em português, linguagem de marcação extensível, é um formato amplamente utilizado para troca de dados, pois possibilita que se mantenha a estrutura dos dados em operações diferentes. O modo como os arquivos XML são construídos permite aos desenvolvedores escrever parte da documentação dentro dos dados, sem interferir na sua leitura. A metalinguagem é usada para criação de outras linguagens e o XML é, portanto, o padrão para troca de dados na Web.

JSON

O formato JSON (do inglês, *JavaScript Object Notation*), pode ser descrito como um padrão de notação de objeto JavaScript, porém sem necessidade exclusiva de utilização do JavaScript, ou seja, seu uso é independe de linguagem de programação. Este formato possui como uma de suas principais características a simplicidade na representação de dados, o que acaba por torná-lo consideravelmente mais rápido em termos de processamento computacional.

Segundo o [W3C 2011], é um formato de arquivo bem fácil de ser interpretado por qualquer linguagem de programação, ou seja, costuma ser mais fácil para os computadores processarem JSON do que outras linguagens, como o XML.

5. SALIC Net

Salic é o acrônimo de Sistema de Apoio às Leis de Incentivo à Cultura, uma plataforma desenvolvida pelo Ministério da Cultura (MinC) para as inscrições de propostas culturais. Desde 2009 o sistema é utilizado para facilitar e agilizar o processo de cadastro, criação de proposta cultural, envio, avaliação e acompanhamento de projetos culturais incentivados pela Lei 8.313 de 23/12/1991, conhecida como *Lei Rouanet*. Em 2013 houve uma reformulação no sistema *Salic WEB* com a implementação do *Novo Salic*. A partir dele, além dos proponentes de projetos passarem a realizar todas as etapas diretamente no sistema, a prestação de contas passou também a ser feita online. Dessa forma, toda a gestão dos processos de projetos culturais passou a ser realizado 100% via sistema, ou seja, excluindo-se qualquer controle paralelo com papéis [Brasil 2014].

A modernização do sistema permitiu que os usuários acompanhassem seus processos de uma maneira bem mais segura. Atividades como solicitações de readequação, prorrogação, fiscalização e comprovações financeiras do projeto são realizadas dentro do próprio sistema. Toda essa infraestrutura acelera o desenvolvimento dos projetos, pois exclui-

se o tempo gasto para envio de documentação ao Ministério da Cultura, além de diminuir os gastos com Correio, evitar erros de procedimento e economizar papel [Brasil 2014].

De acordo com informações do Novo Salic disponíveis em `https://github.com/culturagovbr/novosalic`, os *stakeholders* do processo são:

- Proponentes - Principal usuário do sistema, propõe projetos culturais.
- Incentivadores - Patrocinadores, financiadores dos projetos, podendo ser pessoa física ou jurídica.
- Fornecedores - Agentes que prestam serviços, materiais e estrutura para os projetos.
- Beneficiários - Após a execução do projeto, o Proponente não pode incorporar o patrimônio adquirido para a implementação do projeto (por exemplo sistemas de som), portanto deve doar para um beneficiário (por exemplo coletivos).
- Parecerista - Perito que acessa o sistema e faz a apreciação dos projetos.
- Conselheiro da CNIC - Comissão Nacional de Incentivo à Cultura - Faz a revalidação dos pareceres e sugere a aprovação ou o indeferimento.

Em 2015, durante a 16^a edição do Fórum Internacional Software Livre (FISL16), o MinC anunciou a disponibilização do código do sistema para sociedade por meio de sua página de projetos Github (`https://github.com/culturagovbr/novosalic`). O principal intuito dessa medida era dinamizar o aperfeiçoamento e aprimoramento do sistema, através da participação voluntária de programadores, em princípio sem custos para o MinC (Carlos Paiva e Rômulo Barbosa, correio eletrônico, 14 de agosto de 2015).

De acordo com o artigo 37 da Constituição Federal de 1998, a administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios deverá atender ao princípio da publicidade dos atos da Administração Pública. Neste contexto, houve a necessidade de tornar esses dados disponíveis para serem consultados pela população, promovendo a transparência dos dados governamentais no que diz respeito aos projetos de apoio à cultura.

Os dados cadastrados no Novo Salic são disponibilizados para o público através de uma plataforma governamental de consulta denominada *Salic Net*. Assim, após as propostas serem cadastradas no Salic, podem ser acompanhadas pelo Salic Net, viabilizando uma base de dados com todos os projetos da Ministério da Cultura desde 2009 a fim de garantir maior transparência dos atos praticados pelo Ministério da Cultura.

Por meio do Salic Net é possível acessar as informações sobre os projetos beneficiados pela Lei Rouanet, por meio de consultas, relatórios e extração de dados, tanto de pessoas físicas quanto jurídicas, possibilitando que o cidadão participe da fiscalização e da avaliação das ações do Ministério da Cultura (Sobre o SalicNet, 2007). Embora o sistema apresente diversas funcionalidades para geração de relatórios, é permitido ao usuário somente visualizar os dados, não sendo possível fazer a manipulação dos mesmos para análises mais aprofundadas. Para cobrir esta lacuna, uma API foi desenvolvida para melhor disponibilização dos dados.

6. Application Programming Interface (API)

Uma API (Application Programming Interface) é um conjunto de padrões e rotinas estabelecidos e adotados por um software para promover acesso às suas funcionalidades

por meio de uma interface padrão. No contexto de uma API REST, os softwares clientes acessam os dados de um software que expõe seus recursos fazendo uso de padrões web. Os objetivos principais de uma aplicação dessa natureza são:

- Maior desacoplamento entre produtores e consumidores dos serviços;
- Possibilidade do uso de CACHE de requisições e respostas;
- Preconização de uma arquitetura de camadas;
- Reforço do uso de uma interface padrão de acesso.

Tipicamente uma API REST transporta os dados por meio de HTTP. Esses dados, por sua vez, são usualmente no formato XML ou JSON, conforme discutido anteriormente.

7. API para abertura de dados de cultura

7.1. Características gerais da arquitetura da API do SALIC

A API do SALIC preconizou seguir os padrões arquiteturais e boas práticas de desenvolvimento adotados pela indústria, e recomendados pela literatura especializada. Em termos práticos, seguem as decisões de projeto e suas respectivas justificativas:

- **Linguagem de programação**: Python, devido à sua eficiência, robustez, versatilidade, facilidade de desenvolvimento e segurança;
- **Framework web**: FLASK *framework* [Ronacher 2015], por apresentar simplicidade, flexibilidade, robustez, comunidade ativa, e boa documentação;
- **SGBD** (Serviço gerenciador de banco de dados). SQLServer.

No que refere-se ao SGBD, é importante destacar que o SQLServer é atualmente o sistema utilizado no processo de gerenciamento de dados do sistema do SALIC, fato este que restringiu o desenvolvimento a se principalmente basear no mesmo. Entretanto, considerando a possibilidade de migração para um sistema distinto, quer seja um sistema proprietário ou um sistema sob licença de software livre, como o PostgreSQL [PostgreSQL 2016], durante o desenvolvimento busco-se tomar como base a construção de uma arquitetura que permitisse, sem muito impacto na API, a substituição do SGBD. Neste sentido, considerando estes aspectos, a API do SALIC buscou explorar a tecnologia ORM (*Object-relational mapping*), a qual trata-se de uma abordagem de desenvolvimento que provê o tratamento das tabelas de bancos de dados através de classes, e seus referidos registros por meio de instâncias [Juneau 2013]. Para o desenvolvimento da API do SALIC, a tecnologia ORM foi implementada através do conjunto de ferramentas disponibilizado pelo SQLAlchemy [SQLAlchemy 2016]. A Figura 2 mostra, em termos gerais, a arquitetura da API do SALIC.

De um modo geral, com base na disponibilização das informações relativas às bases de dados de produção do SALIC, discutidas na seção anterior, assim como os esquemas relacionados às funções desenvolvidas para a API do SALIC, a Figura 14 apresenta o domínio da informação resultante, o qual permite compreender de forma mais simples o arranjo de dados e a relação entre os mesmos.

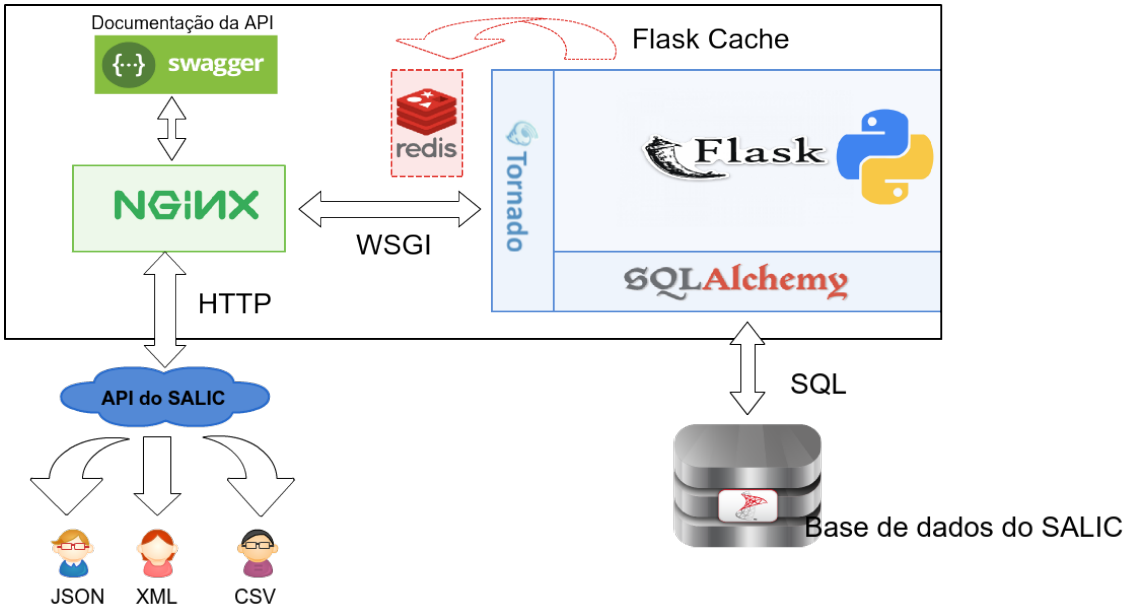


Figura 2. Visão geral da arquitetura da API do SALIC.

7.2. Documentação

A documentação de um sistema pode ser considerada como uma de seus pontos de maior importância, uma vez que, através deste produto é possível melhorar de forma considerável a usabilidade e a exploração das ferramentas disponaíveis por um determinado sistema. Neste sentido, para o uso correto e efetivo de uma API, é imprescindível a construção de um documentação de uso que seja capaz de fornecer informações suficientes e detalhadas a cerca de seus métodos, consultas, disponibilização de resultados, entre outros aspectos.

No refere-se à construção da documentação de usuário da API do SALIC, optou-se por seguir o padrão de documentação Swagger [SmartBear 2016]. Este padrão trata-se de um *framework* que possui como uma de suas bases produzir, através de um modelo específico, uma documentação completa e concisa acerca de todos os recursos disponíveis com suas respectivas estruturas e campos de dados, parâmetros de consultas, formatos suportados e tipos de resposta. Neste sentido, foi gerada uma documentação interativa por meio de uma página web (`http://hmg.api.salic.cultura.gov.br/doc/`), onde é possível tanto explorar as funcionalidades da API de forma simples e direta, quando explorá-la para fins de teste e compreensão, sem necessidade de ferramentas externas, como um cliente já implementado ou um software de terceiro, como é o caso da ferramenta de linha de comando *curl*[2].

7.3. Segurança

A aplicação foi desenvolvida tendo como um de seus principais requisitos manter a segurança e integridade dos dados que serão disponibilizados. Este é um ponto essencial, uma vez que, a API do SALIC atua diretamente com dados que envolvem pessoas físicas e jurídicas, os quais devem ser tanto resguardados de alterações não autorizadas, quanto, em alguns casos, não devem ser disponibilizados publicamente.

[2]Curl é uma ferramenta para tratamento de URL's e transferência de dados através de linha de comando.

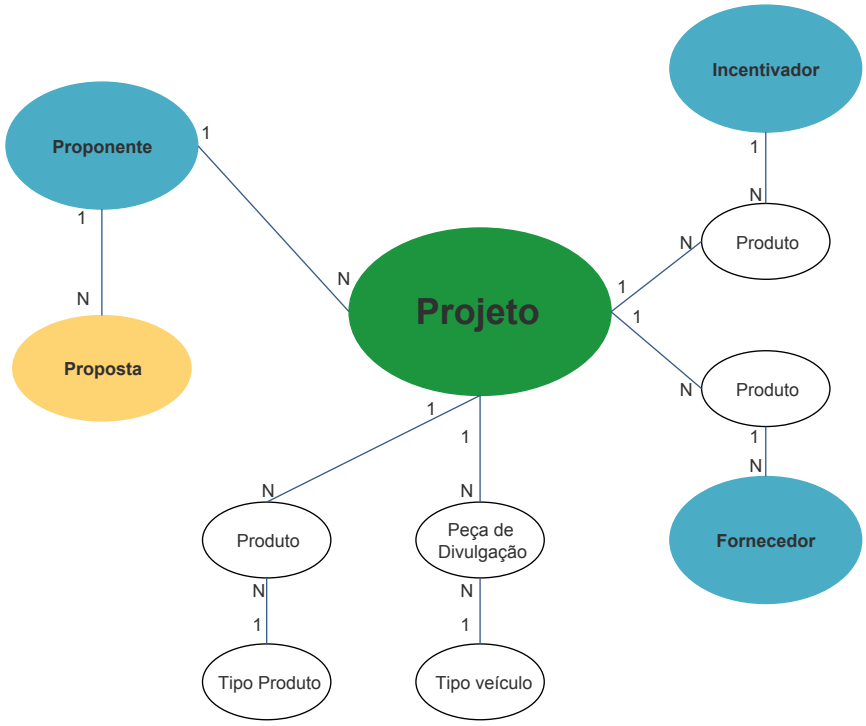


Figura 3. Domínio da Informação da API do SALIC.

Considerando estes fatores, foi feito um trabalho de curadoria para identificação dos dados que poderiam estar disponíveis através da API, onde, através de sanitização dos dados, foram excluídos do escopo dos métodos desenvolvidos dados mais sensíveis, como telefones ou endereços residenciais. Além disso, é importante destacar que a API também conta com um sistema que limita o acesso baseado no endereço IP de origem. Desse modo, um usuário só pode fazer um certo número de requisições diárias, evitando assim possíveis ataques de disponibilidade.

Para além dos requisitos e restrições de segurança considerados, um sistema de autenticação está em desenvolvimento, onde o objetivo é tornar o sistema ainda mais seguro com a identificação de quem acessa o sistema, além de possibilitar políticas de uso individualizadas. Como exemplo, um dado usuário poderia ter um número de acessos diários diferente do padrão que é imposto aos usuários comuns.

Outro aspecto considerado pela API do SALIC, no que refere-se à disponibilidade e não oneração do SGBD de produção do próprio SALIC, é a implementação de um sistema de cache customizado. Essa estratégia teve como objetivo minimizar o acesso direto à base de dados, já que ela já é usada por vários outros serviços que compõem o sistema SALIC. O cache tem como *backend* o banco de dados em memória Redis [Redis Labs 2016], um sistema flexível, rápido e robusto.

7.4. Formatos de resposta

Uma das principais características de uma API para abertura de dados é a disponibilização dos dados em formatos estruturados, os quais permitam a sua reutilização para os mais variados propósitos. Neste sentido, a API do SALIC permite o retorno de requisições

nos formatos JSON, XML e CSV. Esses três formatos compreendem e cobrem o maioria dos usos de um usuário convencional, consumidor de dados dessa característica. Além disso, é importante destacar que existe uma considerável variedade de bibliotecas em diversas linguagens que facilitam consumir dados em JSON e XML, e, por outro lado, o formato CSV é popularmente usado para se trabalhar diretamente com os dados por meio de planilhas eletrônicas

8. Conclusão

Este artigo apresentou os aspectos gerais do desenvolvimento de uma API para abertura de dados de um sistema de gestão de projetos de cultura do Ministério da Cultura do Brasil, com enfoque voltado à questões de caráter mais relacionado à tecnologia da informação. Os resultados obtidos mostraram que este tipo de abordagem, isto é, a disponibilização de dados através de APIs de dados abertos, pode ser considerada uma alternativa consideravelmente interessante no que refere-se ao atendimento à transparência por parte dos órgãos governamentais, assim como também o que diz respeito às possibilidades de reutilização de dados por parte da sociedade. Apesar disso, é importante destacar que um estudo considerando os vários aspectos da abertura de dados, desde a prontidão da organização até fatores de utilização, deverá ser realizado para permitir uma análise mais abrangente e precisa.

Referências

Agune, R. M., Filho, A. S. G., and Bolliger, S. P. (2010). Governo aberto sp: disponibilização de bases de dados e informações em formato aberto. In *Anais do III Congresso Consad de Gestão Pública*, São Paulo.

Brasil, P. (2014). Novo salic é modernizado e vai facilitar acompanhamento de projetos.

Choo, C. W. (2003). *A organização do conhecimento*. Editora Senac, São Paulo.

de Souza Pires, J. C. and Macêdo, K. B. (2006). Cultura organizacional em organizações públicas no brasil. *RAP*, 40(1):81–105.

Juneau, J. (2013). *Java EE 7 Recipes: A Problem-Solution Approach*, chapter Object-Relational Mapping, pages 369–408. Berkely, CA, USA, 1 edition.

Open Knowledge Foundation (2011). *Open Definition*. Open Knowledge Foundation.

PostgreSQL (2016). Postgresql global development group. Acessed; 2016-10-12.

Redis Labs (2016). Redis. Acessed: 2016-10-12.

Ronacher, A. (2015). Flask framework. Acessed: 2016-10-12.

SmartBear (2016). Swagger. Acessed: 2016-10-12.

SQLAlchemy (2016). The database tookkit for python. Acessed: 2016-10-12.

W3C (2011). *Manual dos Dados Abertos: Governo*. W3C.

Proposta de modelo para inclusão da Governança da Informação e da TI na perspectiva da Governança Corporativa

Mauro Cesar Bernardes[1,3], Marino Hilário Catarino[2,3]
[1]Centro Universitário Estácio Radial de São Paulo, São Paulo – SP
[2]Centro Universitário FIEO, São Paulo – SP
[3]Universidade de São Paulo, São Paulo - SP
mcesar.bernardes@gmail.com, marinohc@gmail.com

Abstract*: This paper describes a proposal for a model to govern corporate information by combining processes, roles, controls and metrics from existing frameworks in order to set information as a valuable business asset. The objective is to provide a common, practical and flexible framework to help organizations develop and implement effective information management programs. As result, a structure to organize Information Governance and IT Governance in the Corporate Governance perspective will be presented.*

Resumo: *Este artigo descreve a proposta de um modelo para a governança de informações corporativas por meio da combinação de processos, papéis, controles e métricas de frameworks existentes, a fim de tratar as informações como um ativo de valor. O objetivo é fornecer um quadro comum, prático e flexível para ajudar as organizações a desenvolver e implantar programas de Governança eficazes e viáveis. Como resultado, será apresentado de uma estrutura posicionamento da Governança da Informação e da Governança de TI na perspectiva da Governança Corporativa.*

1. Introdução

Extrair o maior valor da Informação para benefício de uma organização não é uma questão de se trabalhar com maior intensidade ou por um tempo maior. É uma questão de identificação e envolvimento das pessoas certas de diferentes áreas nas decisões sobre como estruturar e fazer uso de Tecnologias da Informação (TI), bem como, de uma concepção de como tomar decisões (nível estratégico) e como colocar em prática tais decisões (nível tático-operacional).

Quando se enfrenta o desafio de administrar a área de TI de uma organização maior, as primeiras questões que surgem são: quais são as metas que se pretende atingir em TI para atender às demandas da área de negócio e qual o ponto de partida. A partir do momento que isto for claro, o próximo passo será definir ações a serem desenvolvidas em um plano de execução para essas ações. Entretanto, a identificação do ponto de partida não é uma atividade trivial, como também não é trivial para a área de TI entender de forma clara as demandas de negócio e traduzi-las em metas de TI.

O ponto de partida deverá incluir o conhecimento da organização, as capacidades da área de TI, o ambiente onde se insere como também a realidade do mercado onde se reside e atua. Isto requer que seja elaborado, com um cuidado quase artesanal, o Plano Estratégico da organização, de onde é derivado o Plano Estratégico de TI e o seu modelo de Governança de TI.

O próximo passo será a definição de um mecanismo de acompanhamento e avaliação da execução do Plano Estratégico, fundamental para se garantir o seu sucesso e viabilizando a realização de ajustes necessários em resposta ao dinamismo do mercado e da própria organização. Considerando esse também um papel da Governança, depara-se com o enorme desafio de especificar direitos decisórios e de um *framework* de responsabilidades para estimular comportamentos desejáveis em todos os níveis da estrutura organizacional, bem como identificar e definir os melhores processos, em um nível tático-operacional, para garantir que a infraestrutura de TI sustente e amplie as estratégias e objetivos da organização. Por sua vez, controlar processos nesse nível é parte integrante do Gerenciamento de Serviços, que apesar de diferenciar-se de Governança de TI, é parte integrante e indispensável.

De forma análoga, uma boa gestão de serviços de TI é fundamental para a Governança da Informação, que por sua vez deverá prover aos gestores dados que confiáveis e úteis durante a tomada de decisões de negócios. Em muitas organizações, as responsabilidades para tarefas de governança da informação são divididas entre as equipes de segurança, armazenamento e banco de dados, responsáveis pelo gerenciamento de serviços de TI. Entretanto, a necessidade de uma abordagem holística para a gestão da informação se torna imprescindível, demandando não apenas visões de gerenciamento, mas, sobretudo de governança. Assim, um primeiro passo deverá ser a compreensão dos domínios de Governança Corporativa, Governança da Informação, Governança de TI e Gerenciamento de serviços de TI.

Em apoio aos responsáveis por esse desafio apresentam-se um conjunto de modelos e metodologias consagradas no mercado, como COBIT, ITIL, *Balanced Scorecard*, Normas Internacionais de Segurança, guia PMBok e avaliação de maturidade. Entretanto, utilizá-los em conjunto, explorando sinergicamente seu potencial, sem se perder em detalhes que poderão desviar do foco principal e, principalmente, identificar o nível organizacional adequado para envolvimento com cada modelo e metodologia tem sido desafiador para diversos profissionais nas mais diversas organizações. O objetivo deste trabalho é apresentar uma análise sobre os temas Governança da Informação, Governança de TI e Gerenciamento de Serviços de TI, diferenciando estes temas e posicionando-os nos níveis hierárquicos organizacionais. Um modelo integrado de Governança de TI que combina potencialidades de *frameworks* consagrados pelo mercado, posicionando-os nos níveis adequados de gestão empresarial, será apresentado. Este modelo guiará ao entendimento das ações necessárias para Gerenciar Serviços de TI e estruturar a Governança da TI e a Governança da Informação no escopo da Governança Corporativa. O resultado da análise, em conjunto com o modelo de governança apresentado, fornecerão subsídios para que as organizações considerem a inclusão da Governança da Informação e da TI na perspectiva da Governança Corporativa.

2. Ponto de Partida: Entendendo Gerenciamento de Serviços de TI, Governança da Informação e Governança de Tecnologia da Informação

Para toda infraestrutura, recursos e informações necessárias para a utilização da TI existem no mercado modelos de Gerenciamento de Serviços de TI e *frameworks* Governança que direcionam, sustentam e controlam administrativa e tecnicamente essa área.

Entretanto, a falta de entendimento claro sobre o escopo da Governança de TI, Governança da Informação e a contribuição das ações de Gerenciamento de serviços de TI neste escopo dificultam a identificação do modelo e da metodologia adequados a serem aplicados em cada nível organizacional, o que resulta muitas vezes na não obtenção do retorno esperado no seu uso. Para a implantação de um modelo de governança de TI, que inclui a adoção de um modelo de gerenciamento de serviços e Governança da Informação, é necessário entender a perspectiva e as necessidades da organização, pois cada contexto tem um modelo específico a ser aplicado. A compreensão das diferenças nestas definições e o correto posicionamento nos níveis organizacionais é o primeiro passo para que ações de Governança de TI, Governança da Informação e de Gestão de Serviços de TI estejam retornando adequadamente o esperado com o investimento nessas ações.

2.1. Governança de TI

A governança de TI surgiu com a intenção de governar estrategicamente a TI, derivando, conforme alguns autores, da governança corporativa. Na definição apresentada pelo ITGI (2004) temos Governança de TI como sendo o conjunto de processos, responsabilidades e ações com objetivos estratégicos afim de atender as necessidades do negócio da organização, ficando a cargo da diretoria e gerência executiva.

Weill e Ross (2006:8) conceituam governança de TI como "a especificação dos direitos decisórios e do framework de responsabilidades para estimular comportamentos desejáveis na utilização da TI". O ITGI (2004) acrescenta que a governança de TI é de responsabilidade da diretoria e gerência executiva da organização, sendo que um fator crucial na governança de TI é conseguir identificar os responsáveis pelas decisões e quem responderá (positiva ou negativamente) por elas. Para Fernandes e Abreu (2008), a Governança de TI busca o compartilhamento de decisões de TI com os demais gestores da organização, assim como estabelece as regras, a organização e os processos que nortearão o uso da TI pelos usuários, departamentos, unidades de negócio, fornecedores e clientes.

O relacionamento entre Governança Corporativa e Governança de TI pode ser representado graficamente como o apresentado na figura 1.

O ITGI (2004) ainda apresenta que o propósito da governança de TI é o de direcionar a TI e as segurar que seu desempenho direcione: o alinhamento da TI com a organização e a realização dos benefícios prometidos; o uso da TI para capacitar a organização para exploração das oportunidades e maximização dos benefícios; o uso responsável dos recursos de TI; e gestão de riscos relacionados a TI.

Um dos aspectos da governança de TI é a questão da política de tomada de decisão. Para Nestor (2001), existem dois lados da governança de TI: o lado normativo, que cria os instrumentos e mecanismos que garantem a formalização de regras e procedimentos operacionais, permitindo o alcance dos objetivos organizacionais. O outro lado, apresentado mais claramente neste artigo, é o comportamental, que estabelece os relacionamentos formais e informais, além de assegurar os direitos decisórios aos grupos ou indivíduos sobre aspectos que envolvem a TI.

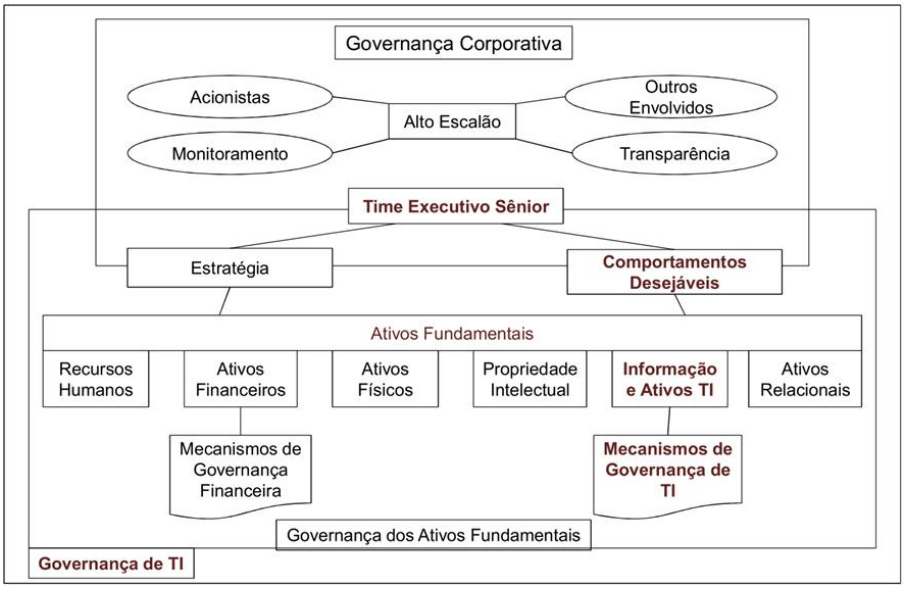


Figura 1: Modelo de Governança Corporativa e Ativos fundamentais. (Fonte: Weill & Ross, 2006)

Assim, de acordo com conceitos de governança de TI, sobretudo no aspecto comportamental, deve-se considerar o entendimento e obtenção de respostas para perguntas como: como as decisões de TI devem ser direcionadas? Quem fará o direcionamento? Quem fornecerá subsídios para as decisões? Quem domina o conhecimento em determinadas áreas que necessitam de decisões? Como controlar e monitorar as decisões tomadas? Quem serão os responsabilizados pelas decisões?

De acordo com Weill e Ross (2006:10), uma governança de TI eficaz deve tratar de três questões referentes à tomada de decisão:

a) Quais decisões devem ser tomadas para garantir a gestão e o uso eficazes de TI?
b) Quem deve tomar essas decisões?
c) Como essas decisões serão tomadas e monitoradas?

Mas para se tomarem decisões é necessário dispor de informações, controles, processos e procedimentos e de um *framework* de responsabilidades para estimular comportamentos desejáveis na utilização da TI (Weill e Ross, 2006). Assim, quanto mais rápida e precisa for a informação, mais eficaz é a gestão e a orientação da área de TI e do negócio para o sucesso.

Todos estes mecanismos estimulam a transparência das instituições para com os envolvidos, evidenciando a aplicação real dos investimentos e comparando o retorno esperado com o alcançado, desejo de todo gestor no momento da quantificação do investimento adequado em TI.

2.2. Governança da Informação

Uma vez identificados os elementos de um modelo de Governança de TI que contemple gerenciamento de serviços, percebe-se atualmente a necessidade de se considerar a Governança da Informação.

A governança da informação fornece normas, diretrizes e controles de responsabilidade para empresas de modo que essas organizações possam obter e garantir a qualidade, valor e *compliance* da informação, sendo estes três as dimensões de destaque da Governança da Informação (Otto, 2011).

A relação existente entre valor da informação e qualidade da informação foi constatada por Zhao (2008). Estas duas dimensões são difíceis de serem mensuradas, por serem subjetivas e definidas conforme as percepções do usuário da informação (Lee et al., 2006).

A qualidade da informação pode ser formalizada em uma empresa através da instituição de um framework de governança da informação (Wende, 2007). Foi verificado que governança da informação pode administrar e aprimorar a qualidade da informação nas organizações (Cheong e Chang, 2007; Panian, 2010).

É difícil governar e se estipular valor de uma informação, já que ela depende muito do usuário (Kooper, Maes e Lindgreen, 2011). Tendo como objetivo o aumento do valor das informações, as organizações precisam administrar seus dados como se fossem ativos empresariais (Panian, 2010).

Por fim, *compliance* é definido como a conformidade da informação com requerimentos legais regulatórios da organização (Datskovsky, 2009). As organizações podem atestar o *compliance* com regulações norteadas por dados, leis e políticas, além do alinhamento com a governança corporativa e governança de TI (Datskovsky 2009; Grimstad, Myrseth, 2011).

2.2.1. Revisão Cronológica para Governança da Informação

Um dos primeiros artigos científicos sobre governança da informação foi publicado em 2004 por Donaldson & Walker (2004), onde o assunto abordado foi a implantação de governança da informação no Serviço Nacional de Saúde na Escócia (Reino Unido). No trabalho foi apresentado o modelo HORUS, para governança da informação, que tinha como finalidade controlar o acesso, gravação e o uso correto e eficiente da informação.

Outros trabalhos também tiveram como base a governança da informação no Serviço Nacional de Saúde do Reino Unido. Tendo como foco os princípios, compromissos e a legislação que os profissionais de saúde deveriam ter para lidar com as informações dos pacientes (Dunnill e Barham, 2007; Barham, 2010).

O modelo *Organizational Knowledge Base* (OKB) foi desenvolvido em um trabalho contendo uma abordagem de governança da informação com elementos de gestão do conhecimento, através de três estudos de casos em empresas suíças (Gianella e Gujer, 2006).

A relação entre governança da informação e segurança da informação foi vista por Williams (2008). Para o autor, governança da informação não tem somente o objetivo de proteção das informações, mas também de garantir a sua disponibilidade quando e onde for solicitada, respeitando a ética e os padrões legais.

Datskovsky (2009) coloca que, fornecer uma maior transparência da gestão da informação, abranger uma regulação da informação em organizações multinacionais atendendo a diferentes regulamentos e leis sobre as informações e por fim a produção

constante de informações dos processos da organização são os três principais objetivos da governança da informação.

A governança de tecnologia de informação deve considerar que a informação é de grande importância nas empresas, e com os desenvolvimentos tecnológicos, o papel e o valor da informação foram alterados (Weill e Ross, 2006). Com grande influência neste artigo, Khatri e Brown (2010) publicaram um trabalho que propõem um modelo de governança da informação. Os autores buscaram um modelo que tivesse um uso funcional, além de auxiliar os pesquisadores quanto à governança de dados. Neste artigo os autores não fazem distinção entre dado e informação.

O artigo de Khatri e Brown (2010) serviu de referência para o artigo publicado por Grimstad e Myrseth (2010), tendo uma visão mais técnica do assunto. Foi apresentada a governança de informação com metadados contextualizando a estrutura da informação em uma análise no setor público.

Rosenbaum (2010) aborda o problema jurídico da informação, publicando um trabalho sobre a gestão em um ambiente de trocas de informações confidenciais, no caso um hospital onde ocorrem trocas de informações cruzadas entre os pacientes, médicos, planos de saúde e fornecedores considerando a responsabilidade da administração e as obrigações legais.

O conceito de que a informação é um fator governado, desvinculado da tecnologia utilizada na organização, oferecendo uma antecipação de desenvolvimentos informativos futuros e também um estimulo para conceitos inovadores em torno do uso da informação foi apresentado por Beijer e Kooper (2010) como uma possibilidade na governança da informação.

Cragg, Caldeira e Ward (2011) estudaram os sistemas de informação em empresas de pequeno e médio porte localizadas em Portugal, Nova Zelândia e Reino Unido. Os autores verificaram que apesar de poucas empresas possuírem uma política formal de gestão da informação, a maioria define responsabilidades e papéis para os gestores e outros funcionários.

Kooper, Maes e Lindgreen (2011) afirmaram que governança da informação permite um melhor governo do uso da informação dentro e fora da organização, para isto utilizam um conceito que contém os atores que produzem, recebem e governam essa informação. Com isto eles se diferem dos outros artigos até então publicados, sendo que a informação é a união de todos os conceitos.

Tallon (2013) conclui que investigações futuras de Tecnologia da Informação devem incluir considerações sobre governança da informação. Segundo ele este é um passo necessário para que os investigadores comecem a avaliar a magnitude do papel de governança da informação nas organizações modernas, pois não somente artefatos físicos de Tecnologia da Informação (*hardware*, *software*) podem ser governados, mas artefatos da informação também devem ser considerados exclusivos e governados.

Governança da informação é ainda pouco explorada e compartilhada entre as empresas, porém a associação com *Big Data* esta ampliando sua implantação. Esta necessidade surge, na maioria dos casos, quando associada a *business intelligence* (de Freitas, 2013). Com isto, novas expectativas em relação aos usos de dados é que com

governança da informação e aplicabilidade de big data, é possível fazer melhores previsões e fazer mais decisões inteligentes.

Para Cleland (2014), há uma tendência global de abertura das informações pelos governos, impulsionada por exigências de maior transparência e prestação de contas e por oportunidades para a inovação no setor público e privado, onde uma ferramenta política importante no setor público tem sido a publicação aberta on-line de dados do governo. Com isto um dos principais problemas que é tratado pela Governança da Informação é o equilíbrio entre a necessidade de proteger determinadas categorias de informações com a necessidade de promover a sua abertura.

2.2.2. Teorias para Governança da Informação

A governança da informação tem como base duas teorias, que tem origem na economia da informação: A teoria descrita por Akerlof (1970) e a teoria descrita por Jensen e Meckling (1976). A teoria de Akerlof (1970) baseia-se na dificuldade em distinguir a qualidade da informação. Esta teoria permite verificar que a existência de informações assimétricas nas organizações produz um prejuízo na qualidade da informação. Este prejuízo indica que o usuário não sabe se a informação recebida é um "limão" ou uma boa informação.

Já a teoria de Jensen e Meckling (1976) tem como foco o relacionamento agente-principal. Esta teoria também busca solucionar os problemas de assimetria de informação e conflitos de interesse entre os membros das organizações (Klann, Beuren e Hein, 2008). A teoria da agência procura melhorar o relacionamento, através de mecanismos de monitoramento do processo, como estruturas de governança, visando conter o problema de agência (Frezatti et al., 2009).

2.3. Gerenciamento de Serviços de TI

O gerenciamento de serviços de TI pode ser conceituado como o conjunto de práticas gerenciais que tem como foco o fornecimento de serviços e produtos de TI a uma organização, ou seja, o gerenciamento das operações de TI.

A Governança de TI, por outro lado, é mais ampla, e se concentra na viabilização e na transformação da TI para atender às necessidades atuais e futuras do negócio (foco interno) e dos clientes (foco externo) (Peterson, 2004). A figura 2 ilustra as orientações da Governança e do Gerenciamento, e seus respectivos horizontes de tempo.

Enquanto a Governança de TI deve preocupar-se com a definição de papéis, responsabilidades, processos, políticas, padrões, diretrizes para o uso adequado dos recursos da TI, planejamento estratégico da TI, identificação e priorização dos projetos relevantes para sua implantação e sustentação, além do controle de investimentos e orçamento, o Gerenciamento de Serviços de TI, por sua vez, deverá ter foco na entrega dos resultados definidos pelas metas de TI ou pelas demandas apresentadas pelas áreas de negócio, garantindo a disponibilidade da operação (o dia a dia).

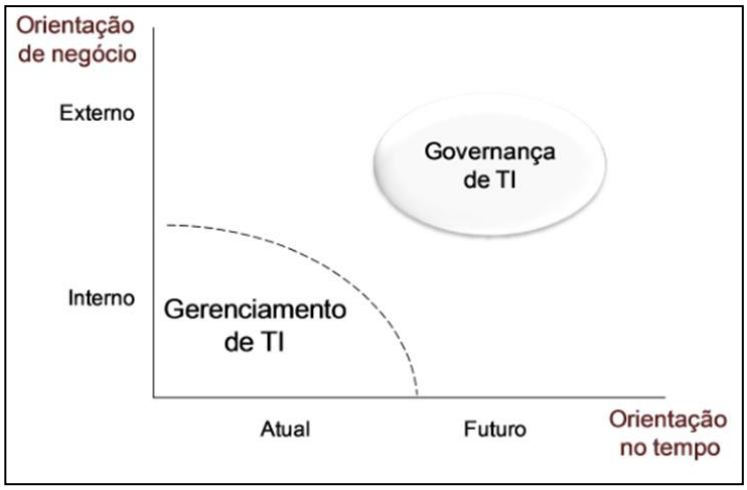


Figura 2: Orientações de Governança e Gerenciamento de Serviços de TI (Fonte: Peterson, 2004)

Portanto, o Gerenciamento de serviços de TI tem como foco eficiência operacional, provisão de serviços e recursos de TI, a definição de "como fazer", enquanto Governança de TI terá como foco no direcionamento do "o que fazer". A correta identificação dos níveis organizacionais envolvidos com a Governança e o Gerenciamento de Serviços de TI deve ser vista como o passo inicial para a identificação e posicionamento de modelos e *frameworks* relacionados em uma organização. A figura 3 apresenta uma visão gráfica para auxílio nessa identificação.

Figura 3: Governança Gerenciamento de Serviços de TI – posicionamento. (Fonte: Bernardes, 2015)

O gerenciamento de serviços de TI deve envolver a definição de um conjunto de processos a serem realizados pelas unidades provedoras de TI, visando ao planejamento e à realização das atividades necessárias ao provimento ou entrega de soluções e serviços de TI a uma organização.

A diferença entre Gestão de serviços de TI e Governança de TI reside também no foco e no *locus* das atividades: enquanto a gestão foca o ambiente interno da organização e é realizada no nível operacional e tático, a Governança de TI congrega o foco interno e externo e deve ser realizado em nível hierárquico superior (estratégico),

de modo a englobar a organização como um todo (Sethibe, Campbell e McDonald, 2007).

2.3. Execução e Controle de Processos no Gerenciamento de Serviços de TI

Considerando que a Governança de TI apresenta às organizações o desafio da cultura de gestão baseada em medições, que considera não apenas aspectos financeiros, mas controle de fatores que apontem para o futuro por meio de indicadores, permitindo uma melhor definição qualitativa quanto ao desempenho da TI no resultado alcançado, torna-se indispensável que o Gerenciamento de Serviços de TI tenha foco a gestão da TI baseada a execução de processos no nível operacional e o respectivo controle desses processos no nível tático. A figura 4 ilustra esse posicionamento.

Figura 4: Gerenciamento de serviços de TI: Execução e Controle de processos. (Fonte: Bernardes, 2015)

Uma vez identificados e definidos os processos prioritários ao bom gerenciamento dos serviços de TI, alcança-se os requisitos essenciais para o atendimento de uma das premissas indispensáveis à Governança de TI: a medição para o controle. O princípio da medição vem ao encontro do monitoramento das atividades e serviços providos por TI para garantir que os processos sejam qualificados e conhecidos em seus detalhes. Quando ocorre tal conhecimento da cadeia de valor de processos, eles podem ser controlados e alinhados de acordo com os objetivos do negócio. Somente a partir desse controle, realizado no nível tático, é possível ocorrer o gerenciamento eficaz, a base para a Governança.

Dessa forma, o gerenciamento de serviços de TI, que deve ter foco na geração eficaz de produtos e serviços de TI e no gerenciamento das operações de TI para uma organização, deve ser considerado parte complementar e indispensável à Governança de TI, sendo que os modelos apresentados para este não poder ser confundidos como *frameworks* de Governança de TI.

2.4. Posicionamento dos diferentes níveis de Governança e Gerenciamento de Serviços de TI na Governança Corporativa

Como resultado da análise dos principais elementos que compõem um modelo de Governança de TI, as ações necessárias ao gerenciamento de serviços de TI e os

principais conceitos de Governança da Informação, observa-se que a Governança Corporativa em uma organização deverá contemplar esses elementos como um modelo de camadas, apresentado na figura 5.

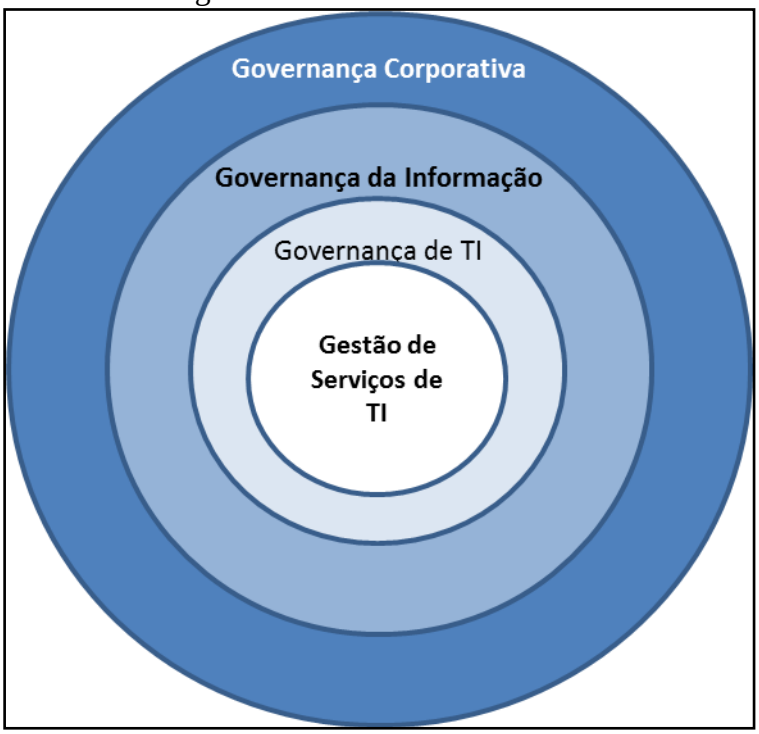


Figura 5: Modelo em camadas para Governança de TI e da Informação no âmbito da Governança Corporativa

Um comparativo dos elementos de Governança Corporativa, da Informação e de TI é apresentado na tabela 1. A compreensão desses elementos direcionará a organização para o desenvolvimento de uma Governança Corporativa que contemple os níveis adequados de Governança e Gerenciamento de Serviços de TI.

Tabela 1: Comparativo entre Governança Corporativa, Governança de TI e Governança da Informação

	Governança Corporativa	Governança de TI	Governança da Informação
Princípios	Direitos dos acionistas;	Alinhamento com governança corporativa;	Alinhamento com governança corporativa;
	Independência;	Fornecer maior valor da TI para a empresa;	Fornecer maior valor da informação para a empresa;
	Responsabilidade e transparência;	Administrar os riscos da TI.	Assegurar a utilização e a administração correta das informações.
	Papéis e responsabilidades da equipe de administração.		
Fatores Motivadores	Crescimento da complexidade na organização;	Os mesmos da governança corporativa;	Os mesmos da governança corporativa;
	Globalização;	Suporte da governança corporativa;	Suporte da governança corporativa;
	Rápido avanço da tecnologia;	Expansão do papel da TI: nas iniciativas de estratégia;	Roubo e invasão de dados;
	Aceleramento da tomada de decisão;	Expansão do papel da TI: no conhecimento administrativo;	Má qualidade da informação;
	Equipes mais proativas;	Expansão do papel da TI: privacidade, segurança e continuidade;	Aumento e maior rapidez do compartilhamento das informações;
	Aumento da competição;	Proliferação de tecnologias do tipo "solução".	Responsabilidades legais das empresas quanto a informação.
	Escândalos recentes.		
Influências da Teoria da Agência	Mecanismos que buscam minimizar:	Mecanismos que buscam minimizar:	Mecanismos que buscam minimizar:
	Custo de agência;	Custo de agência;	Custo de agência;
	Assimetria de informações.	Assimetria de informações.	Assimetria nas informações.

2.5. Considerações finais

A distinção entre Governança de TI, Governança da Informação e Gerenciamento de Serviços de TI torna-se fundamental no processo de inserção destes conceitos na Governança Corporativa. Enquanto Governança de TI e Gerenciamento de serviços de TI lidam com os ativos de TI da empresa, a Governança da Informação envolve os ativos da informação. A diferença principal reside no fato de os ativos de TI referirem-se às tecnologias que auxiliam e apoiam a automação de tarefas bem definidas, enquanto os ativos de informação (ou de dados) são definidos como fatos que têm valor em potencial e que são documentos. A Governança da Informação deve direcionar os mecanismos necessários à Governança de TI, direcionando suas ações. Por sua vez, a Governança de TI deve incluir os mecanismos necessários para o correto gerenciamento dos serviços de TI, o que torna as ações cada vez mais granulares nesta abordagem.

Com essa análise, a Governança Corporativa deve ser vista como o nível mais elevado de Governança em uma organização e um aspecto chave direcionador da Governança da Informação e da Governança de TI. Processos de Governança da Informação devem ser de mais alto nível que a governança de TI que, por sua vez, são de mais alto nível e direcionadores dos detalhes necessários ao gerenciamento dos serviços de TI. A abordagem de Governança da Informação não deve ter como foco os detalhes da TI ou dos dados capturados, mas no controle das informações geradas pela infraestrutura de TI existente.

Ao abordar os conceitos relacionados à Governança Corporativa, à Governança da Informação, à Governança de TI e ao gerenciamento de serviços de TI, este artigo apresenta elementos que permitem lidar com a grande confusão existente a respeito dos termos relacionados à Governança na área de TI.

Com a visão do modelo apresentado neste artigo, o sucesso da governança da informação dependerá de avaliação contínua das políticas e de sua adaptação como prioridades de negócios e condições para a evolução corporativa. Assim como uma governança corporativa estratégia pode render vantagens competitivas, um controle eficaz de informações pode transformar essas informações em um gerador mais consistente do valor comercial desejado.

5- Referências

AKERLOF, George A. (1970). *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*. In: The Quarterly Journal of Economics, Vol. 84, No. 3. (Aug., 1970), pp. 488-500.

BARHAM, Chris (2010). *Confidentiality and security of information*. Anaesthesia and Intensive Care Medicine, v. 11, n. 12, 2010.

BEIJER, Peter; KOOPER, Michiel. (2010) *Information governance: Beyond risk and compliance*. PrimaVeraWorking Papers,2010.

BERNARDES, Mauro César. *Estruturação de um modelo de Governaça de TI combinando metodologias, modelos e ferramentas em diferentes níveis organizacionais.* In: Workshop Hispano-Brasileño de Gobernanza Empresarial de

Tecnologías de la Información, 2015, Santander: Editora de la Universidad de Cantabria, 2015. v. 1. p. 63-78.

CHEONG, Lai Kuan; CHANG, Vanessa. (2007). *The need for data governance: A case study*. In: Australasian Conference on Information Systems (ACIS), 18, Toowommba. Proceedings… Toowommba: 2007.

CLELAND, Brian, GALBRAITH, Brendan, QUINN, Barry e HUMPHREYS, Paul (2014). *Information Governance Modularity in Open Data*. In: Proceedings Of The 9th European Conference On Innovation And Entrepreneurship (ECIE 2014), pp. 108-117.

CRAGG, Paul; CALDEIRA, Mário; WARD, John. (2011). *Organizational information systems competences in small and medium-sized enterprises*. Information & Management, v. 48, 2011.

DATSKOVSKY, Galina. (2009). *Information governance*. In: LAMM, Jacob. Under control: Governance across the enterprise. New York: Apress, 2009.

DAVENPORT, Thomas H., PRUSAK, Laurence e STRONG, Bruce. (2008). *Business Insight (A Special Report): Organization; Putting Ideas to Work: Knowledge management can make a difference - But it needs to be more pragmatic*. In: Wall Street Journal, p. R11, 2008.

de FREITAS, Patricia Alves, dos REIS, Everson Andrade, MICHEL, Wanderson Senra, GRONOVICZ, Mauro Edson e RODRIGUES, Marcio Alexandre de Macedo (2013). *Information Governance, Big Data and Data Quality*. In: CSE '13 Proceedings of the 2013 IEEE 16th International Conference on Computational Science and Engineering, Pages 1142-1143.

DONALDSON, Alistair; WALKER, Phil. (2004). *Information governance – a view from the NHS*. International Journal of Medical Informatics, v. 73, 2004.

DUNNILL, Richard; BARHAM, Chris. (2007). *Confidentiality and security of information*. Anaesthesia and Intensive Care Medicine, v. 8, n. 12, 2007.

FERNANDES, Aguinaldo A.; ABREU, Vladimir F. *Implantando a governança de TI: da estratégia à gestão dos processos e serviços*. 2. ed. Rio de Janeiro: Brasport, 2008.

FREZATTI, Fábio et al. (2009). *Controle gerencial: Uma abordagem da contabilidade gerencial no contexto econômico, comportamental e sociológico*. São Paulo: Atlas, 2009.

GRIMSTAD, Terje; MYRSETH, Per. (2010). *Information governance as a basis for cross-sector e- Services in public administration*. In: E-CHALLENGES E-2010 CONFERENCE, 2010, Warsaw. Proceedings… Warsaw, 2010.

ITGI. *IT Governance Institute. Board briefing on IT Governance*. 2 nd ed. 2004. Disponível em: <www.itgi.org>. Acesso em: 28 nov. 2011.

JENSEN, Michael e MECKLING, William H. (1976). *Theory of the firm: Managerial behavior, agency costs and ownership structure*. Journal of Financial Economics, 1976, vol. 3, issue 4, pages 305-360.

KHATRI, Vijay; BROWN, Carol. (2010). *Designing data governance*. Communications of the ACM, v. 35, n. 1, Jan 2010.

KLANN, Roberto Carlos; BEUREN, Ilse Maria; HEIN, Nelson. (2008). *Impacto das diferenças entre as normas contábeis brasileiras e americanas nos indicadores de desempenho de empresas brasileiras participantes da governança corporativa*. In: Encontro Anual Da ANPAD, 32, 2008, Rio de Janeiro. Anais.... Rio de Janeiro, Associação Nacional dos Cursos de Pós- Graduação em Administração, 2008.

KOOPER, M. N.; MAES, R.; LINDGREEN, E.E.O. Roos. (2011). *On the governance of information: Introducing a new concept of governance to support the management of information*. International Journal of Information Management, v. 31, 2011.

LEE, Yang W. et al. (2006). *Journey to data quality*. Cambridge: MIT Press, 2006.

NESTOR, Stilpon. *International efforts to improve corporate governance: why and how*. Paris: Organization for Economic Co-operation and Development (OECD), 2001.

OTTO, Boris. (2011). *Data governance*. Business & Information Systems Engineering, v. 4, 2011.

PANIAN, Zeljko. (2010). *Some practical experiences in data governance*. World Academy of Science, Engineering and Technology, v. 62, 2010.

PETERSON, R. *Integration strategies and tactis for information technology governance*. In: Grembergen, W. V. (ED). Strategies for implementing information technology governance. Hershey: Idea Group, 2004. p. 37-80.

ROSENBAUM, Sara. (2010). *Data governance and stewardship: Designing data stewardship entities and advancing data access*. Health Services Research, v. 45, n. 5, Out 2010.

SETHIBE, T.; CAMPBELL, J.; McDONALD, C. *IT Governance in Public and Private Sector Organizations. Examinig the Differences and Defining Future Research Directions*. In: Australasian Conference on Information System IT Governance in the Public Sector, Toowoomba, 2007.

TALLON, Paul P., RAMIREZ, Ronald V. e SHORT, James E (2013). *The Information Artifact in IT Governance: Toward a Theory of Information Governance*. In: Journal of Management Information Systems, 30(3), pp.141-178.

Van GREMBERGEN, Win.; HAES, Steven. (2009). *Enterprise Governance of Information Technology*. New York: Springer, 2009.

WEILL, Peter; ROSS, Jeanne W. (2006). *Governança de tecnologia da informação: Como as empresas com melhor desempenho administram os direitos decisórios de TI na busca por resultados superiores*. São Paulo: McBooks, 2006.

WENDE, Kristin. (2007). *A model for data governance: Organising accountabilities for data quality management*. In: Australasian Conference On Information Systems, 18, 2007, Toowoomba. Proceedings. Toowoomba: 2007.

ZHAO, Yuyang et al. (2008). *High value information in engineering organisations*. International Journal of Information Management, v. 28, 2008

Ambiente Integrado de Desenvolvimento para o sistema de Programação Lógica Indutiva Aleph

Jonathan Christian F. Rocha[1], **Rogerio Salvini**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

jonathancfr@gmail.com, rogeriosalvini@inf.ufg.br

Abstract. *Inductive Logic Programming (ILP) combines machine learning with logic programming. The advantage of this approach compared to traditional machine learning algorithms is that it is possible to generate relational and interpretable classifiers. Among the various ILP systems, one of the most complete is the Aleph system. However, unlike traditional machine learning methods, that have several frameworks for the implementation of experiments, the Aleph system lacks in such a feature. The goal of this paper is to present an Integrated Development Environment (IDE) so that users of any field can make research and experiments in ILP using Aleph system intuitively and in a practical way.*

Resumo. *Programação Lógica Indutiva (ILP, na sigla em inglês) combina aprendizado de máquina com programação lógica. A vantagem desta abordagem em relação aos algoritmos de aprendizado de máquina tradicionais é a possibilidade de gerar classificadores relacionais e interpretáveis. Dentre os diversos sistemas ILP existentes, um dos mais completos é o sistema Aleph. Entretanto, ao contrário dos métodos tradicionais de aprendizado de máquina, que possuem diversos frameworks para a execução de experimentos, o sistema Aleph é carente de tal facilidade. O objetivo deste trabalho é apresentar um Ambiente Integrado de Desenvolvimento (IDE), de forma que usuários de qualquer área possam fazer pesquisas e experimentos em ILP usando o sistema Aleph de forma intuitiva e prática.*

1. Introdução

Com a grande quantidade de dados sendo gerados de forma extremamente rápida nos dias atuais, e com avanço na capacidade de armazenamento destes dados, o acúmulo de informações torna-se impossível de ser analisado por seres humanos, sendo necessária uma forma automática de análise de dados e extração de conhecimento.

As ferramentas e técnicas empregadas para análise automática e inteligente destes imensos repositórios são os objetos tratados pelo campo de pesquisa da descoberta de conhecimento em bancos de dados (KDD, do inglês *Knowledge Discovery in Databases*). A mineração de dados é a etapa em KDD responsável pela seleção dos métodos a serem utilizados para localizar padrões nos dados, seguida da efetiva busca por padrões de interesse numa forma particular de representação, juntamente com a busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão [Fayyad et al. 1996].

Em contrapartida, o aprendizado de máquina é uma área de pesquisa bem estabelecida da Ciência da Computação que estuda métodos computacionais para adquirir

conhecimento de forma automática, induzido por exemplos. Algoritmos de aprendizado de máquina são amplamente utilizados na mineração de dados para extração de conhecimento e reconhecimento de padrões.

Dentre as técnicas tradicionais e mais conhecidas de aprendizado de máquina podemos citar: Redes Neuronais Artificiais, Árvores de Decisão, Máquinas de Vetor Suporte (SVM) e métodos bayesianos, como as Redes Bayesianas. Estas técnicas são referidas como ”atributo-valor”ou ”proposicionais”, pois utilizam o formato atributo-valor como linguagem de representação dos objetos. Neste formato os objetos (ou exemplos) são descritos em termos de atributos e valores por meio de um vetor (*vetor de características*) contendo valores para os atributos de um determinado exemplo e um rótulo que atribui uma classe ao exemplo. Portanto, todo o conjunto de dados é representado por uma única tabela onde cada linha representa um objeto e as colunas os seus atributos.

Por outro lado, sistemas de aprendizado relacional, como a Programação Lógica Indutiva, utilizam uma descrição relacional baseada na linguagem de primeira ordem da Lógica Clássica [Lavrač and Džeroski 2001], onde os objetos são descritos em termos de seus componentes e relações entre estes componentes. Estes sistemas têm sido usados com sucesso na extração de modelos relacionais de dados, ou seja, dados armazenados em múltiplas tabelas. Entre suas principais características estão:

- Alta expressividade para representar conceitos;
- Capacidade de representação do conhecimento do domínio;

A tarefa de um sistema de Programação Lógica Indutiva é produzir um classificador representado em Lógica de Primeira Ordem, dados exemplos positivos, exemplos negativos, a descrição destes exemplos e um conjunto de restrições que define a forma como um classificador deve ser construído.

O objetivo deste trabalho é apresentar um Ambiente Integrado de Desenvolvimento (IDE) que foi desenvolvido para o sistema de Programação Lógica Indutiva Aleph. Esta IDE oferece ferramentas de edição de texto, manipulação de arquivos, execução de experimentos, dentre outras facilidades, que auxiliarão usuários a usar o sistema de forma simplificada e ajudará a difundir a Programação Lógica Indutiva dentro da área de aprendizado de máquina.

Este artigo está organizado da seguinte forma: na Seção 2 serão apresentados os conceitos básicos sobre Programação em Lógica Indutiva e será discutido em mais detalhes o sistema Aleph; na Seção 3 serão abordados alguns *frameworks* existentes para algoritmos de Aprendizado de Máquina; Na Seção 4 será apresentado o AlephIDE, assim como sua arquitetura e funcionamento; por fim, na Seção 5 serão apresentados alguns trabalhos futuros e considerações finais.

2. Programação Lógica Indutiva

Programação Lógica Indutiva (ILP, do inglês *Inductive Logic Programming*) combina métodos de aprendizado indutivo (aprendizado a partir de exemplos) com o poder de representação da Lógica de Primeira Ordem, concentrando-se em particular na representação de hipóteses como programas lógicos [Muggleton and de Raedt 1994]. Dentre as vantagens da ILP estão:

- Oferece uma abordagem rigorosa para o problema geral de aprendizado indutivo baseado em conhecimento;
- Oferece algoritmos completos para induzir teorias gerais de primeira ordem a partir de exemplos, o que pode, portanto, aprender com sucesso em domínios onde algoritmos baseados em atributos são difíceis de aplicar. Lógica de Primeira Ordem é uma linguagem apropriada para descrever relacionamentos;
- É possível adicionar informações adicionais sobre o domínio além da descrição dos exemplos;
- Produz hipóteses que são facilmente compreendidas (facilmente interpretáveis pelas pessoas).

Estas características são de grande apelo para pesquisadores das mais variadas áreas do conhecimento, e sistemas em ILP têm sido usados com sucesso na extração de modelos de dados relacionais. Na verdade, ILP já vem sendo usada há mais de uma década para construir modelos preditivos de dados em diversos domínios que incluem Bioinformática [King et al. 1992], Engenharia [Dolsak and Muggleton 1991], Processamento de Linguagem Natural [Zelle and Mooney 1993], Meio Ambiente [Džeroski et al. 1995], Engenharia de Software [Bratko and Grobelnik 1993], Aprendizado de Padrões e Link Discovery [Davis et al. 2007, Mooney et al. 2004] e Alias Identification [Davis et al. 2005].

ILP utiliza dados de um conjunto de exemplos de treinamento $E = E^+ \cup E^-$, onde E^+ e E^- são, respectivamente, conjuntos de exemplos positivos e negativos, e dados do conhecimento do domínio B (também chamado de *background knowledge*), representados como conjuntos finitos de cláusulas. A tarefa da ILP é achar uma hipótese h que satisfaça as restrições de linguagem (*language bias*) e que, quando conjunta com B, implique logicamente todos os exemplos positivos ($B \wedge H \models E^+$) mas nenhum dos exemplos negativos ($B \wedge H \neq E^-$) [Lavrač and Džeroski 2001]. Para lidar com questões do mundo real, contudo, frequentemente relaxam-se os requisitos, tal que h precise somente implicar significativamente em mais exemplos positivos do que exemplos negativos.

A maioria dos sistemas ILP usa um algoritmo que iterativamente examina cláusulas candidatas (espaço de busca) a fim de encontrar cláusulas "boas", que modelem os dados da melhor forma possível de acordo com critérios pré-determinados (tipo de função de avaliação das cláusulas, tamanho máximo da cláusula, quantidade de memória disponível e tempo de computação, entre outros), e formar uma *teoria*, que é um conjunto de cláusulas que será a hipótese gerada pelo classificador.

2.1. Aleph

O Aleph [Srinivasan 2004] é um sistema ILP *open-source*, escrito completamente em Prolog, desenvolvido na Universidade de Oxford. Ele possui uma linguagem poderosa que permite representar expressões complexas e incorporar novos conhecimentos do domínio com facilidade. Além dessas características, comuns aos sistemas ILP, o Aleph permite a alteração de vários parâmetros que permitem ao usuário configurar as estratégias de busca a fim de encontrar a melhor hipótese de acordo com o domínio em estudo.

Como entrada, o Aleph toma informação do conhecimento do domínio na forma de predicados da Lógica, uma lista de tipos e modos declarando como estes predicados podem ser encadeados, e a designação de um *predicado alvo*, que representa a hipótese

a ser aprendida. Também são necessárias listas de exemplos positivos e negativos do predicado alvo.

O algoritmo base do Aleph segue um procedimento que pode ser descrito em quatro passos [Srinivasan 2004]:

1. Seleciona um exemplo positivo para ser generalizado. Se não existir mais exemplos, pare;
2. Constrói uma cláusula mais específica baseada no exemplo selecionado anteriormente e das restrições da linguagem. Essa é normalmente uma cláusula com muitos literais e é chamada *"bottom clause"*. Esse passo é chamado de saturação;
3. Procura por uma cláusula mais geral que a *"bottom clause"*. Esse passo é chamado de redução;
4. A cláusula com melhor pontuação é adicionada a teoria, e todos os exemplos redundantes são removidos. Retorne para o passo 1;

Para a execução do Aleph são requeridos os seguintes arquivos:

- *file.b*: Contém informações sobre o conhecimento do domínio, o qual deve ser escrito na forma de cláusulas Prolog que codifiquem as informações relevantes ao contexto. Este arquivo também contém as restrições de buscas, restrição de linguagem, restrições de tipos e restrições de parâmetros. No quadro 1 abaixo é mostrado um exemplo de um trecho deste tipo de arquivo.

```
:- set(i,2).

:- modeh(1,eastbound(+train)).
:- modeb(1,short(+car)).
:- modeb(1,closed(+car)).
:- determination(eastbound/1,short/1).
:- determination(eastbound/1,closed/1).

%descricao do exemplo: train 1
short(car_12).
closed(car_12).
long(car_11).
```

1. Listagem de um trecho do arquivo *file.b*

- *file.f*: Contém os exemplos positivos do conceito a ser aprendido pelo Aleph. Estes exemplos são representados como fatos Prolog sobre o domínio, como apresentado no quadro 2 abaixo.

```
eastbound(east1).
eastbound(east2).
eastbound(east3).
```

2. Listagem de um trecho do arquivo *file.f*

- *file.n*: Contém os exemplos negativos. Assim como os exemplos positivos, os exemplos negativos são fatos Prolog sobre o domínio, como no quadro 3 abaixo.

```
eastbound(west1).
eastbound(west2).
eastbound(west3).
```

3. Listagem de um trecho do arquivo *file.n*

Uma vez que os arquivos acima estejam criados é possível fazer a execução do Aleph. O resultado é uma lista de cláusulas geradas, juntamente com as melhores cláusulas encontradas, e as respectivas coberturas de exemplos positivos e negativos, conforme mostrado no quadro 4 abaixo.

```
eastbound(A) :- has_car(A,B).
[5/5]
eastbound(A) :- has_car(A,B), short(B).
[5/5]
eastbound(A) :- has_car(A,B), open_car(B).
[5/5]
eastbound(A) :- has_car(A,B), shape(B,rectangle).
[5/5]
[theory]
[Rule 1] [Pos cover = 5 Neg cover = 0]
eastbound(A) :- has_car(A,B), short(B), closed(B).
[pos-neg] [5]
```

4. Listagem de um trecho da saída de execução do Aleph

O uso mais avançado do Aleph permite ainda alteração em cada um dos passos descritos no algoritmo base, possibilitando assim uma busca mais eficiente pela melhor hipótese. É possível determinar uma variedade de restrições no espaço das possíveis hipóteses a serem aprendidas, bem como na busca realizada nesse espaço. Essas restrições podem ser feitas através da modificação de parâmetros.

Dentre os parâmetros mais utilizados estão:

- *clauselength*: define um limite superior para o número de literais em uma cláusula aceitável
- *evalfn*: define a função de avaliação na busca de uma clásula aceitável
- *i*: define um limite superior para o encadeamento de novas variáveis em uma cláusula aceitável
- *minacc*: define um limite inferior na precisão mínima de uma cláusula aceitável
- *minpos*: define um limite inferior no número de exemplos positivos a serem cobertos por uma cláusula aceitável
- *nodes*: define um limite superior para os nós a serem explorados na busca de uma cláusula aceitável
- *noise*: define um limite superior para o número de exemplos negativos autorizado a ser coberto por uma cláusula aceitável

Atualmente, todo o processo para executar experimentos de aprendizado usando o Aleph é feito manualmente. Além da criação dos arquivos necessários, o usuário precisa

digitar comandos através da janela de terminal do sistema operacional. Em geral, os passos a serem seguidos são:

1. Abrir uma janela de terminal (ou *prompt*) do sistema operacional
2. Executar o interpretador Prolog
3. Carregar (consultar) a implementação prolog do Aleph (*aleph.pl*)
4. Fazer a leitura dos arquivos do experimento (*file.b*, *file.f* e *file.n*)
5. Entrar com o comando para gerar uma nova teoria
6. Entrar com o comando para salvar a teoria gerada

Uma dificuldade ainda maior, principalmente para pesquisadores que não são da área da computação, é a realização de experimentos utilizando validação cruzada, a fim de avaliar a teoria gerada com relação à sua capacidade de generalização, ou seja, de classificar novos exemplos. A validação cruzada consiste em particionar o conjunto de dados D em n subconjuntos e então executar o algoritmo desejado n vezes, cada vez usando um conjunto de treinamento diferente $(D - D_i)$ e validando os resultados com o subconjunto $D_i, 1 \leq i \leq n$, [Blockeel and Struyf 2002]. Neste caso, o usuário necessita criar as partições do conjunto D, seguir os passos para a execução do Aleph em cada partição, registrar os resultados das partições e calcular o desempenho final segundo a(s) métrica(s) de interesse.

Na Seção 4 apresentamos um ambiente gráfico que permite ao usuário abstrair comandos manuais, simplificar a execução de experimentos e analisar os resultados usando o Aleph.

3. *Frameworks* para algoritmos de Aprendizado de Máquina

Nesta seção apresentamos alguns *frameworks* populares cujas funcionalidades foram estudas e serviram como base para o desenvolvimento do sistema proposto neste trabalho.

Um *framework* é um conjunto de classes que colaboram para realizar uma responsabilidade para um domínio de um subsistema da aplicação [Fayad and Schmidt 1997]. Os frameworks podem ser classificados como sendo verticais ou horizontais. Os frameworks verticais ou frameworks especialistas, são desenvolvidos com base na experiência obtida em um domínio específico. Os frameworks horizontais são um caso mais genérico, no qual a aplicação que o utilizará pode compreender qualquer âmbito de uso [Sommerville 2007]. Portanto, um framework para algoritmos de aprendizado de máquina consiste em métodos de aprendizado de máquinas propriamente ditos, com o objetivo de atacar um problema específico. Dentre os frameworks para o referido contexto, os mais conhecidos são:

- WEKA[1]: Desenvolvido pela Waikato University, na Nova Zelândia, o WEKA é um projeto *open-source* que foi escrito na linguagem de programação Java. O mesmo contém uma coleção de algoritmos de Aprendizado de Máquina e Mineração de Dados, o qual possui ferramentas para pré-processamento de dados, classificadores, regressão, clusterização, associação de regra e visualização [Hall et al. 2009].

[1]Disponível em: `http://www.cs.waikato.ac.nz/ml/weka/downloading.html`

- Orange[2]: É uma ferramenta *open-source* de Aprendizado de Máquina e Mineração de Dados implementada usando *scripts* Python e C++, desenvolvida na University of Ljubljana. A biblioteca de Orange oferece diversas funcionalidades, tais como: pré-processamento e gerenciamento de dados, classificação, regressão, associação, clusterização e projeção [Demšar et al. 2013].
- KEEL[3]: É uma ferramenta *open-source* escrita na linguagem de programação Java projetada para acessar algoritmos evolucionários para problemas de Mineração de Dados. Inclui métodos de regressão, classificação e aprendizado não supervisionado. Fornece diferentes técnicas para pré-processamento de dados permitindo análises completas de qualquer modelo de aprendizado. [Alcalá-Fdez et al. 2008].
- Rapid Miner[4]: Desenvolvido pela Rapid-I, na University of Dortmund, Alemanha. O RapidMiner é uma ferramenta *open-source* e oferece funções específicas para Mineração de Dados, tais como árvores de decisão, clusterização e associação de regras. Apesar de fornecer algumas funcionalidades gratuitas, o Rapid Miner possui versões empresárias que oferecem uma quantidade maior de recursos [GmbH 2010].

É importante salientar que os *frameworks* supra citados, bem como outros, tratam apenas com algoritmos de aprendizado proposicionais, ou seja, cujos dados são baseados em tabelas de dados atributo-valor e nenhum possui implementações de algoritmos ILP ou aprendizado relacional.

4. AlephIDE

Neste trabalho, apresentamos um ambiente integrado de desenvolvimento chamado AlephIDE. Este sistema foi desenvolvido principalmente pela carência de um *framework* para o Aleph, que atualmente é executado por comandos no terminal/*prompt* do sistema operacional, sendo necessário algum conhecimento, não só de como os arquivos devem ser dispostos, mas também de quais comandos e como estes devem ser executados para realizar determinada tarefa. O AlephIDE permite ao usuário abstrair comandos manuais e facilita a execução de experimentos.

Vale ressaltar que nenhuma funcionalidade do sistema foi retirada sendo que usuários mais avançados serão capazes de usufruir de todos os recursos do Aleph, pois além de permitir a criação e acesso de novos projetos, a IDE permite também importar, de forma transparente, arquivos do Aleph já existentes criados anteriormente pelos usuários.

4.1. Requisitos de software

Para o desenvolvimento deste sistema utilizou-se a linguagem de programação Java, a qual foi escolhida pelas facilidades de criação e manipulação de interfaces gráficas, assim como a sintaxe e o amplo manual disponível. Portanto, devido ao uso dessa linguagem, um dos requerimentos do sistema é a presença da Máquina Virtual Java (*Java Virtual Machine*, JVM[5]). A JVM é uma máquina de computação abstrata semelhante a uma máquina

[2]Disponível em: `http://orange.biolab.si/download/`

[3]Disponível em: `http://sci2s.ugr.es/keel/download.php`

[4]Disponível em: `https://my.rapidminer.com/nexus/account/index.html#downloads`

[5]Disponínel em: `https://www.java.com/en/download/manual.jsp`

real, possui um conjunto de instruções e manipula regiões de memória, e é capaz de executar programas e *scripts*, independente do hardware e sistema operacional.

Além da JVM, são também necessários os seguintes softwares:

- Arquivo ”aleph.pl”[6]: implementação do Aleph
- YAP Prolog[7] (*Yet Another Prolog*): um interpretador de alta performance Prolog, desenvolvido na Universidade do Porto

4.2. Arquitetura

A interface da IDE foi feita na língua inglesa com intuito de abranger uma gama maior de usuários no mundo todo. A arquitetura do sistema está organizada através dos seguintes pacotes:

1. *gui*: É responsável por gerir toda a interface presente no contexto do sistema. A classe *“MainScreen”* centraliza todas as funcionalidades, sendo de responsabilidade da mesma chamar as outras classes a fim de executar tarefas mais específicas;
2. *control*: Gerencia todas as estruturas de dados, manipulações de arquivos e projetos;
3. *induce*: Cuida da criação e comunicação entre processos (AlephIDE e YAP Prolog), assim como gera os comandos necessários para o funcionamento do sistema;

Esses pacotes trocam informações entre si, que por sua vez, trocam informações entre as classes neles existentes. A Figura 1 mostra como as classes estão dispostas.

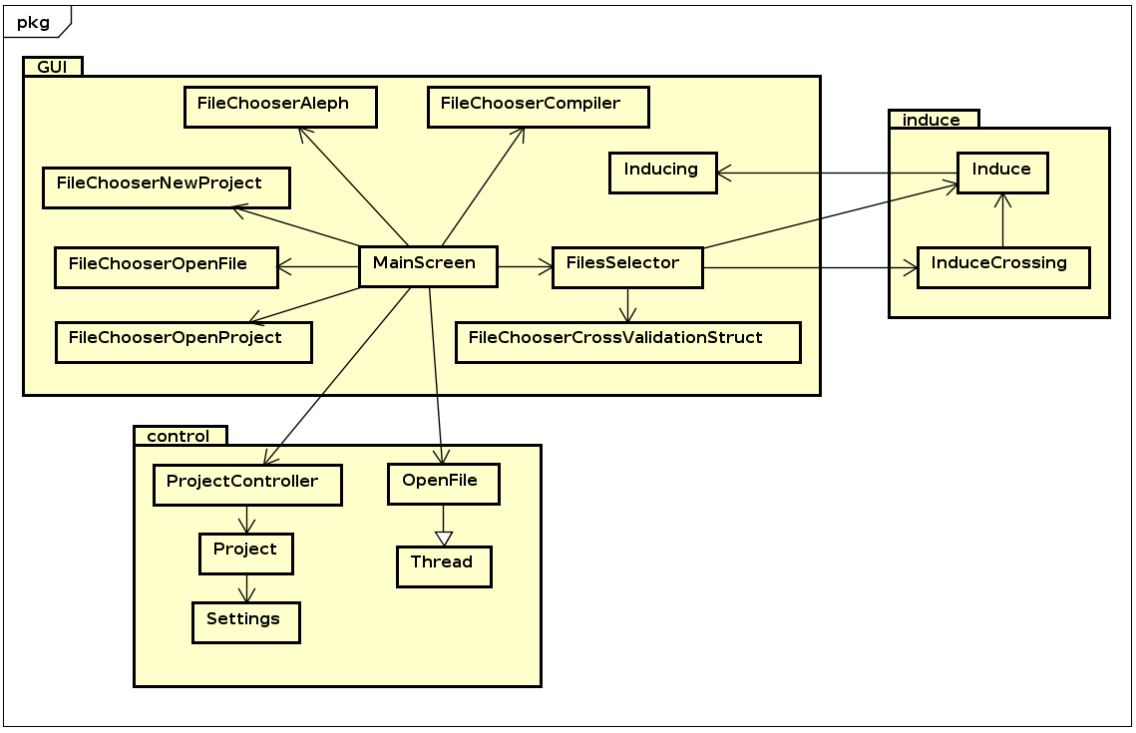


Figura 1. Diagrama de Classes para o AlephIDE

[6]Disponínel em: `http://www.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/aleph.pl`

[7]Disponínel em: `https://www.dcc.fc.up.pt/~vsc/Yap/downloads.html`

No desenvolvimento do AlephIDE, foi necessário estabelecer um *link* entre o processo Java e o processo do interpretador Prolog (YAP), a classe *"Induce"* é responsável por gerenciar tal comunicação. A mesma utiliza uma primitiva Java de criação de subprocessos (*ProcessBuilder*), que através do método *start( )* é capaz de executar um conjunto de comandos. O YAP Prolog por sua vez, recebe como argumento, todas as instruções que lhe foram passadas. Isso é feito pelo comando apresentado no quadro 5.

```
ProcessBuilder().command(compiler, "-l", alephPath)
.redirectInput(run).redirectOutput(out).start();
```

5. Comando para execução do interpretador YAP Prolog

O argumento *compiler* é uma variável que contém o caminho do intepretador Yap Prolog, o *-l* carrega o arquivo Prolog que implementa o sistema Aleph (aleph.pl), que por sua vez tem seu caminho contido na variável *alephPath*. O método *redirectInput(run)* redireciona a entrada padrão do interpretador Prolog, para o arquivo *run*, que contém todos os comandos referentes ao projeto em si, tais como: o arquivo de configuração, os arquivos de treino e teste, e os comandos para gerar e salvar a teoria. Por último o método *redirectOutput(out)* redireciona a saída padrão para o arquivo *out*.

No referido contexto, o intepretador Prolog está apto a ser executado, nesse momento o processo Java que o chamou, fica em estado de espera até que o processamento termine. A Figura 2 mostra como esse processo funciona:

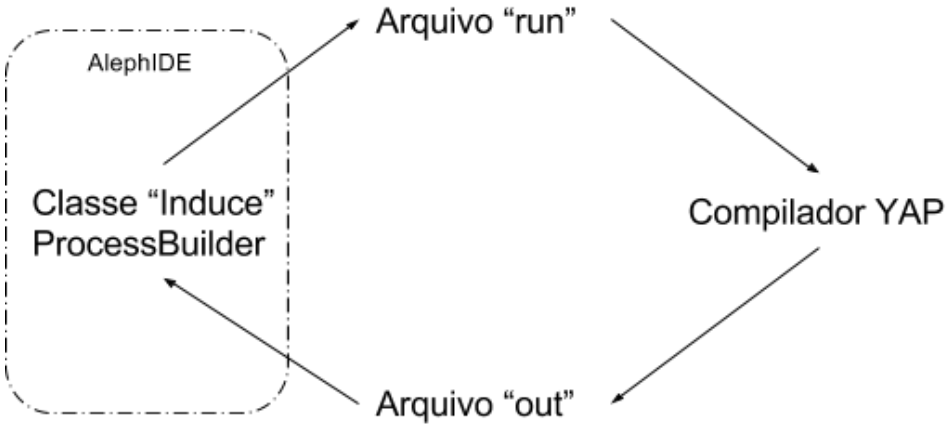


Figura 2. Interação entre o processo Java do AlephIDE e o processo do interpretador Prolog

4.3. Interface com o usuário

Nesta seção serão apresentadas as principais funcionalidades do AlephIDE que já estão em funcionamento. A Figura 3 mostra a tela principal do sistema.

A interface do AlephIDE é constituída por:

1. Um menu, o qual é dividido em cinco sessões: *"File"*, *"Edit"*, *"Run"*, *"System"* e *"Help"*. A sessão *"File"* permite ao usuário criar, abrir, fechar e excluir projetos, assim como criar, abrir, salvar e excluir arquivos. A sessão *"Edit"* disponibiliza ferramentas para edição de texto, como *"Undo"*, *"Redo"* e *"Find"*. A

sessão *"Run"* fornece acesso às janelas de execução e configuração de parâmetros do Aleph para o projeto ativo. A *"System"* permite configurar o caminho do interpretador Prolog e da implementação do Aleph. E por último a sessão *"Help"*, que fornece algumas informações sobre o sistema.

2. Uma barra de ferramentas, a qual possui atalhos para as opções mais comuns.
3. Na lateral esquerda existe um navegador hierárquico de diretórios, dando ao usuário liberdade de manipular os arquivos presentes em cada projeto.
4. Na região central está localizado o editor de texto, o mesmo possui um mecanismo de abas, permitindo abrir vários arquivos ao mesmo tempo.
5. Na região inferior existe uma janela de mensagens do sistema.

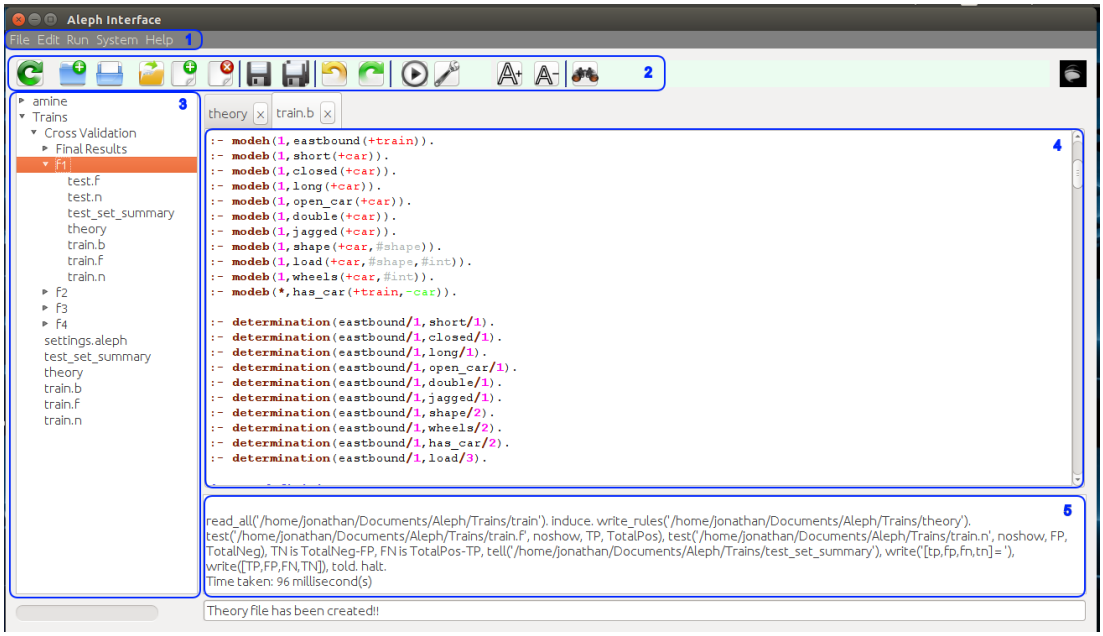


Figura 3. Tela inicial do AlephIDE

Na Figura 4 é apresentada a tela para a execução de experimentos. O usuário pode escolher como será feita a avaliação do classificador por meio das seguintes opções:

1. *"Use Training Set"*: A teoria gerada é avaliada no próprio conjunto de treinamento.
2. *"Supplied Training Set"*: O usuário escolhe os arquivos de teste para avaliar a teoria gerada.
3. *"CrossValidation"*: O usuário escolhe a quantidade de partições (*folds*) para fazer a validação cruzada. O sistema cria os arquivos de treinamento e os arquivos de teste a partir do conjunto original. Estes arquivos são armazenados em subdiretórios do projeto e podem ser usados posteriormente, fora da interface. São oferecidas três formas para a criação dos arquivos nos subdiretórios:
    - *"Round-Robin"*: Os exemplos são distribuídos de forma sequencial nos arquivos dos subdiretórios;
    - *"Random"*: Os exemplos são distribuídos de forma aleatória nos arquivos dos subdiretórios;

- *"Choose Struct"*: O usuário escolhe uma estrutura de diretórios já existente com os arquivos de treinamento e teste. Esta opção é especialmente útil quando o usuário deseja comparar os resultados do Aleph com o de outros algoritmos usando os mesmos exemplos de treinamento e teste, ou quando o usuário importa um projeto já existente com os arquivos gerados fora da interface.

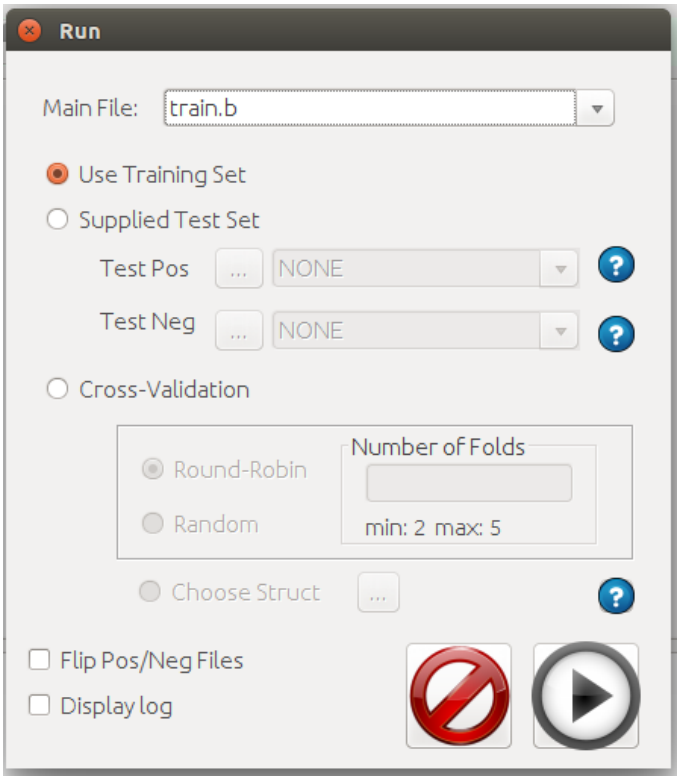


Figura 4. Tela para a execução de experimentos

Conforme descrito na Seção 2.1, o algoritmo do Aleph gera cláusulas a partir dos exemplos positivos. Através da opção *"Flip Pos/Neg Files"* é possível inverter as classes de exemplos para que seja gerada uma nova teoria a partir dos exemplos da outra classe.

Ao final de uma execução, o sistema fica encarregado de mostrar a teoria gerada (Figura 5) e também cria um arquivo chamado *"theory"* que fica armazenado no diretório do projeto.

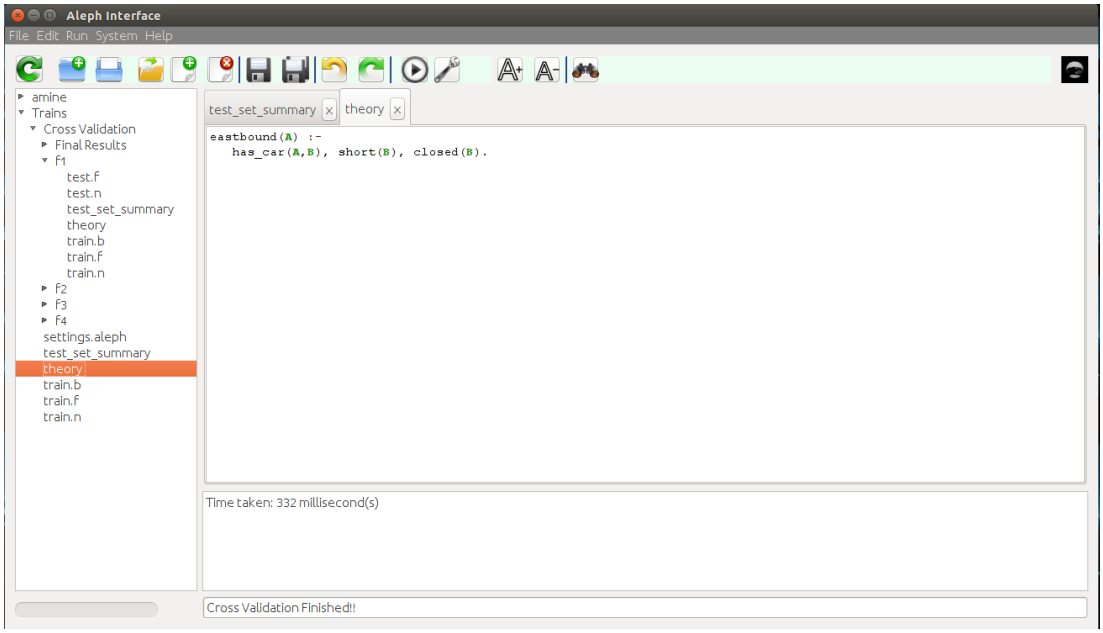


Figura 5. Tela da teoria gerada após a indução

O sistema mostra também os resultados alcançados sobre os conjuntos de teste. É apresentada a matriz de confusão, na forma de um vetor, com a quantidade de exemplos positivos-verdadeiros (*tp*), falso-positivos (*fp*), falso-negativos (*fn*) e negativos-verdadeiros (*tn*). A partir destes valores são calculadas várias medidas de desempenho, como: acurácia (*Accuracy*), acurácia balanceada (*Balanced Accuracy*), sensibilidade (*Recall*), precisão (*Precision*), especificidade (*Specificity*) e medida-F (*F-measure*)[Salvini 2008]. É mostrado também o tempo total decorrido da execução do experimento (Figura 6).

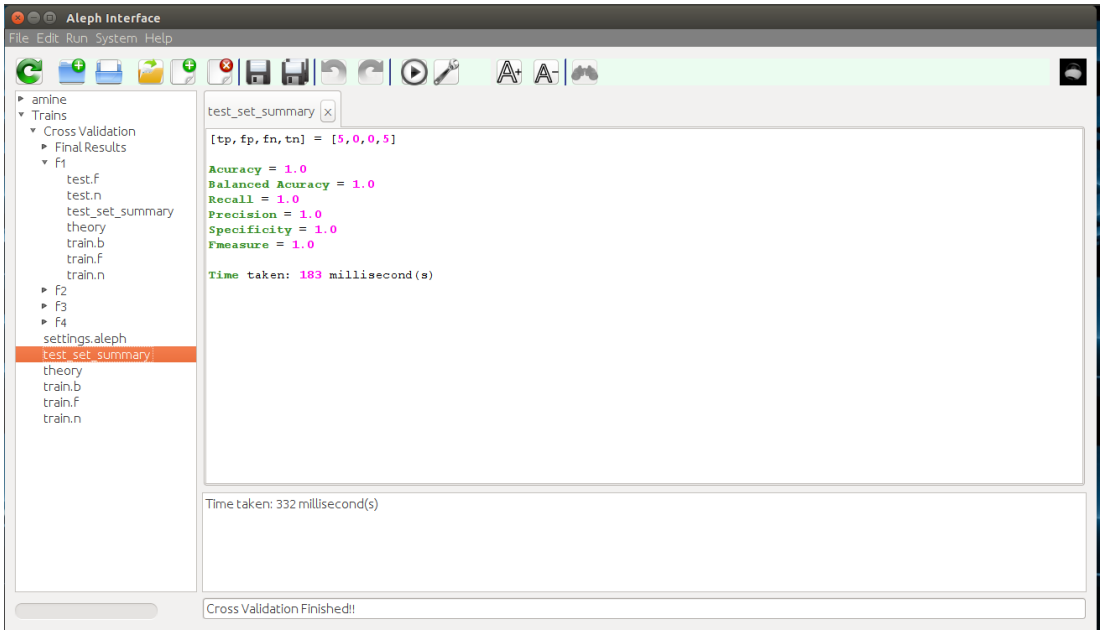


Figura 6. Tela das métricas geradas após a indução

5. Conclusão

Algoritmos de aprendizado de máquina têm sido amplamente utilizados na descoberta de padrões e extração de conhecimento em bases de dados cada vez maiores e diversificadas.

Sistemas de aprendizado relacional, como a Programação Lógica Indutiva (ILP), conseguem resolver problemas de aprendizado multi-relacional utilizando conhecimento do domínio, além dos exemplos, para gerar hipóteses, e produzem classificadores que são de fácil entendimento por especialistas de qualquer área, característica fundamental quando o objetivo é obter novos conhecimentos acerca de um domínio específico.

O Aleph é um sistema ILP robusto, desenvolvido totalmente em Prolog, que permite ainda que o usuário ajuste diversos parâmetros a fim de melhorar a busca pela melhor teoria (classificador). Entretanto, a manipulação dos arquivos necessários para utilizar o Aleph, bem como a necessidade da inserção manual de comandos através de janelas de terminais do sistema operacional podem acabar limitando o seu uso para um público mais amplo, principalmente para usuários que hoje em dia estão acostumados a manusear sistemas através de interfaces gráficas.

Dessa forma, apresentamos neste trabalho uma interface que intermedeie a interação do usuário com o Aleph, a qual fornece um ambiente de desenvolvimento intuitivo. Os comandos são substituídos por botões, e os arquivos são dispostos em um modelo hierárquico cuja edição pode ser feita na própria interface. O usuário pode mudar configurações de parâmetros através de uma tela própria para isso. Os experimentos também são realizados através de uma tela específica onde, por exemplo, fazer o processo de validação cruzada passa a ser uma tarefa trivial.

Trabalhos futuros no AlephIDE incluem o aprimoramento do sistema para importação de dados diretamente de arquivos no formato *csv* (*comma-separated values*) e a inclusão de novas funcionalidades que permitam que o usuário possa fazer experimentos em ILP mais sofisticados.

Referências

Alcalá-Fdez, J., Sánchez, L., García, S., del Jesus, M. J., Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., Fernández, J. C., and Herrera, F. (2008). Keel: a software tool to assess evolutionary algorithms for data mining problems. Springer Verlag.

Blockeel, H. and Struyf, J. (2002). Efficient algorithms for decision tree cross-validation. *Journal of Machine Learning Research*, pages 621–622.

Bratko, I. and Grobelnik, M. (1993). Inductive learning applied to program construction and verification. pages 279–292.

Davis, J., Dutra, I. C., Page, D., and Costa, V. S. (2005). Establishing identity equivalence in multi-relational domains. In *Proceedings of the International Conference on Intelligence Analysis, McLean, VA, USA, May 2–6*. Awarded best technical paper.

Davis, J., Page, D., Burnside, E., Dutra, I., Ramakrishnan, R., Shavlik, J., and Costa, V. S. (2007). *Introduction to Statistical Relational Learning*, chapter Learning a New View of a Database: With an Application in Mammography. MIT Press.

Demšar, J., Curk, T., Erjavec, A., Črt Gorup, Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., and Zupan, B. (2013). Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, 14:2349–2353.

Dolsak, B. and Muggleton, S. (1991). The application of ILP to finite element mesh design. In Muggleton, S., editor, *Proceedings of the First International Workshop on Inductive Logic Programming (ILP-91)*, pages 225–242.

Džeroski, S., Dehaspe, L., Ruck, B., and Walley, W. (1995). Classification of river water quality data using machine learning. In *Proceedings of the 5th International Conference on the Development and Application of Computer Techniques to Environmental Studies*.

Fayad, M. and Schmidt, D. C. (1997). Object-oriented application frameworks. *Communications of the ACM*, 40:32–38.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *American Association for Artificial Intelligence*, pages 37–54.

GmbH, R.-I. (2010). Rapidminer benutzerhandbuch. Disponível em: `http://docs.rapidminer.com/`, Último acesso em Outubro de 2016.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). In *The WEKA Data Mining Software: An Update*, volume 11. SIGKDD Explorations.

King, R., Muggleton, S., and Sternberg, M. (1992). Predicting protein secondary structure using inductive logic programming. *Protein Engineering*, 5:647–657.

Lavrač, N. and Džeroski, S. (2001). *Relational Data Mining*. Springer Verlag.

Mooney, R. J., Melville, P., Tang, R. L., Shavlik, J., Dutra, I., Page, D., and Costa, V. S. (2004). *Data Mining: Next Generation Challenges and Future Directions*, chapter Relational Data Mining with Inductive Logic Programming for Link Discovery, pages 239–254. AAAI Press.

Muggleton, S. and de Raedt, L. (1994). Inductive Logic Programming: Theory and Methods. *Journal of Logic Programming*, 19(20):629–679.

Salvini, R. (2008). *Uma nova metodologia para melhoramento de classificadores baseados em ILP*. PhD thesis, COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação.

Sommerville, I. (2007). *Engenharia de Software*. Pearson, São Paulo, 8th edition.

Srinivasan, A. (2004). *The Aleph Manual*. Disponível em: `http://www.cs.ox.ac.uk/activities/machinelearning/Aleph/`, Último acesso em Outubro de 2016. Oxford University.

Zelle, J. and Mooney, R. (1993). Learning semantic grammars with constructive inductive logic programming. pages 817–822, Washington, D.C. AAAI Press/MIT Press.

Um levantamento sobre pesquisas em Sistemas de Informação Inteligentes no período 2008-2015 no contexto do Simpósio Brasileiro de Sistemas de Informação

Maickon Assunção da Silva, Renato de Freitas Bulcão Neto

Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 - 74.690-900 – Goiânia – GO – Brasil

maickon.assuncao@gmail.com, renato@inf.ufg.br

Abstract*. Due to technological development and the possibility of applying in different areas of knowledge, intelligent systems are increasingly being studied, developed and integrated into contemporary society, from the daily tasks of ordinary people to solve complex problems. This paper presents an investigation about the state of the art of researches on Intelligent Information Systems in the Brazilian context. Thus, a systematic review of published papers on Brazilian Symposium on Information System in the last eight years in order to identify gaps and trends of research in that matter was held.*

Resumo*. Devido ao desenvolvimento tecnológico e a possibilidade de aplicação em diferentes áreas do conhecimento, os sistemas inteligentes estão, cada vez mais, sendo estudados, desenvolvidos e inseridos na sociedade contemporânea, desde as tarefas cotidianas das pessoas comuns, até a resolução de problemas complexos. Este artigo apresenta uma investigação sobre o estado da arte das pesquisas em Sistemas de Informação Inteligentes no contexto brasileiro. Para isso, foi realizada uma revisão sistemática da literatura sobre os trabalhos publicados no Simpósio Brasileiro de Sistemas de Informação nos últimos oito anos com o objetivo de identificar lacunas e tendências de pesquisa no referido assunto.*

1. Introdução

Sistemas de Informação Inteligentes são sistemas capazes de pensar e agir de forma racional, ou como seres humanos [Russel and Norvig 2003]. Eles incorporam no seu modo de operar aspectos relacionados com a dimensão inteligente do comportamento humano, tais como raciocínio, aprendizagem, evolução, adaptação, autonomia, interação social ou proatividade. Eles não são capazes somente por armazenar e manipular dados, mas também por adquirir, representar e manipular conhecimento, com capacidade para deduzir novos conhecimentos. Ou seja, inferir novas relações sobre fatos e conceitos a partir do conhecimento existente para a solução de problemas complexos.

A cada ano, o avanço tecnológico em ritmo acelerado e o número cada vez maior de pessoas conectadas à Internet a partir de diferentes tipos de aparelhos e gerando um aumento exponencial na geração de dados tem viabilizado e tornado cada vez mais necessário o desenvolvimento de sistemas inteligentes, que se tornaram fundamentais em praticamente todos os sistemas de informação e passaram a fazer parte do cotidiano das pessoas. O grande potencial da aplicação destes sistemas em diferentes áreas do conhecimento e setores da sociedade motiva o desenvolvimento cada vez mais interdisciplinar dos sistemas inteligentes, com contribuições de diversas áreas, como Engenharia, Ciência da Computação e Ciências Sociais e Humanas.

O objetivo deste trabalho é analisar o estado da arte das pesquisas em Sistemas de Informação Inteligentes no contexto brasileiro, a partir da análise dos trabalhos completos produzidos no tema citado nos últimos oito anos. O objetivo é identificar as principais teorias de inteligência artificial utilizadas, do foco destes trabalhos, suas principais contribuições, os domínios de aplicação e as principais questões de pesquisa em aberto descritas nos trabalhos.

A metodologia utilizada para se alcançar este objetivo foi um processo de revisão sistemática da literatura [Kitchenham 2004] que consiste em um estudo de um conjunto de artigos retornados de bases de pesquisa guiado por um protocolo de revisão sistemática. Os artigos utilizados como referencial para o desenvolvimento desta revisão foram pesquisas relacionadas com Sistemas de Informação Inteligentes e publicadas no Simpósio Brasileiro de Sistemas de Informação (SBSI), por se tratar da maior conferência sobre a área de Sistemas de Informação no Brasil, todas armazenadas na Biblioteca Digital Brasileira de Computação (BDBComp) que atende aos critérios da fonte de pesquisa definidos no protocolo da revisão.

2. Metodologia

Uma revisão sistemática é uma forma de pesquisa que utiliza como fonte de dados a literatura sobre determinado tema [Sampaio 2007]. É uma forma de identificar, avaliar e interpretar todas as pesquisas que foram encontradas e definidas como relevantes para uma determinada questão de pesquisa, área temática, ou fenômeno de interesse. Neste trabalho, o objetivo da revisão sistemática foi levantar os dados sobre o estado da arte das pesquisas em Sistemas de Informação Inteligentes no Brasil, o qual teve como referencial o trabalho de [Kitchenham 2004], que sugere o seguinte processo:

- **Planejamento:** fase em que é elaborado o protocolo que norteia todo o processo de revisão, definindo as questões de pesquisa, os tipos de trabalhos a serem pesquisados, os filtros a serem aplicados nas buscas pelos trabalhos, as fontes de pesquisa que reúnem as publicações sobre o tema definido, dentre outros[1];
- **Execução:** execução das strings de busca nas bibliotecas digitais selecionadas para a aquisição dos artigos e documentação dos resultados obtidos em planilha;
- **Seleção:** leitura de metadados como título e resumo dos artigos coletados na etapa anterior, executando os critérios de exclusão definidos no protocolo e identificando na planilha os trabalhos que foram excluídos;
- **Extração:** leitura na íntegra dos artigos selecionados na etapa anterior e execução dos critérios de exclusão, restando apenas os artigos relevantes para a obtenção dos dados que responderão às questões definidas no protocolo;
- **Análise e Conclusão:** produção de gráficos, dados estatísticos e conclusões acerca dos dados dos artigos considerados relevantes na etapa anterior.

2.1 Fase de Execução

Nesta fase do processo de revisão sistemática, os trabalhos a serem utilizados no processo de revisão foram buscados na fonte de pesquisa definida no protocolo. Diferentemente do método adotado por [Kitchenham 2004] para a realização da fase de

[1] Para obter detalhes sobre o protocolo de revisão utilizado neste trabalho, acesse: drive.google.com/open?id=1KOqf7b3LyTb-UYsWM1k-y9PyvtzeBYDXZex9EUDKhiA

execução, que consiste em executar strings de busca em fontes de pesquisa definidas no protocolo, neste trabalho os artigos utilizados foram obtidos através da colaboração dos coordenadores (*chairs*) de cada edição do evento SBSI. Estes encaminharam planilhas com os trabalhos completos e aceitos no SBSI em cada uma das edições do evento ocorridas de 2008 a 2015, edições cujos trabalhos publicados estão acessíveis pela BDBComp, com seus respectivos dados compilados do sistema JEMS. Vale ressaltar que esses trabalhos foram classificados com o tópico de interesse relacionado a Sistemas de Informação Inteligentes pelos próprios autores dos trabalhos durante o ato da submissão no JEMS. Juntas, estas planilhas somaram 113 artigos.

Ao invés de strings de busca definidas, foram utilizados metadados como autores, título e ano de publicação de todos os 113 artigos compilados pelo JEMS e registrados nas planilhas enviadas pelos *chairs* para a busca pelos trabalhos na base BDBComp. As buscas retornaram 106 artigos (94%), sendo os outros 7 artigos (6%) não encontrados por motivos de resultados retornados com *links* quebrados, que encaminhavam para algum trabalho diferente do procurado, ou buscas que retornaram resultados vazios a partir dos dados fornecidos de algum artigo. Os 106 artigos retornados (Gráfico 1) tiveram seus dados documentados em uma planilha eletrônica, e seguiram para a próxima fase desta revisão.

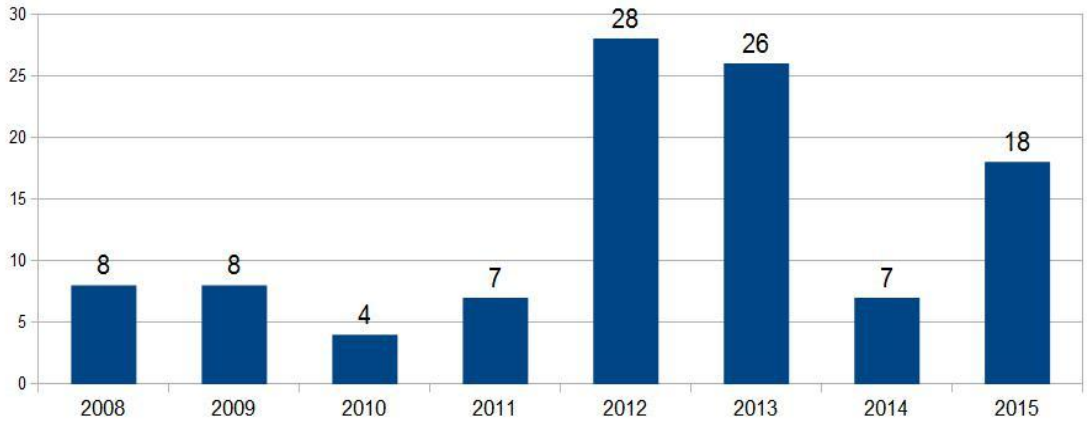


Gráfico 1. Fase de Execução: artigos retornados da base BDBComp.

2.2 Fase de Seleção

Após as buscas pelos trabalhos na fonte de pesquisa BDBComp, a partir da leitura do título e resumo dos artigos retornados foram aplicados os critérios de exclusão definidos no protocolo, para selecionar os artigos relevantes para este trabalho. Dos 106 artigos extraídos, apenas 3 (2 de 2008 e 1 de 2009) foram eliminados nesta etapa. Este resultado se deu pelo fato de que, nesta etapa, os artigos não foram lidos na íntegra, dificultando as eliminações pelos demais critérios. Os 103 artigos restantes seguiram para a próxima fase desta revisão sistemática.

2.3 Fase de Extração

Para a execução da fase de extração todos os 103 artigos foram lidos na íntegra confrontados com os critérios de exclusão. Após a execução dos critérios de exclusão nesta fase, foram eliminados mais 24 trabalhos. O Gráfico 2 relaciona esses artigos eliminados aos respectivos critérios de exclusão.

Gráfico 2. Fase de Extração: critérios de exclusão.

Após a leitura dos artigos foi possível eliminar os trabalhos sob os demais critérios de exclusão. Dos artigos eliminados nesta fase, a maioria (58%) foi excluída por não ser trabalho completo. Segundo os critérios do evento SBSI, artigos completos são trabalhos de pesquisas já concluídas ou em estágio avançado de desenvolvimento devendo possuir de 7 a 12 páginas (incluindo as figuras, tabelas, diagramas, bibliografia e anexos). Os artigos curtos são pesquisas em andamento limitados a 6 páginas. Os outros 10 (42%) artigos eliminados nesta fase não estavam relacionados com o tema Sistemas de Informação Inteligentes.

Portanto, após esta fase é possível concluir que dos 106 artigos extraídos inicialmente da base BDBComp, 34 deles (30%) foram eliminados sob a execução dos critérios de exclusão definidos, restando um total de 79 trabalhos relevantes sobre o tema Sistemas de Informação Inteligentes, publicados entre 2008 e 2015, seguindo o protocolo de revisão sistemática definido. Esses trabalhos foram utilizados para responder às questões de pesquisa definidas no protocolo, apresentadas a seguir.

3. Resultados da Revisão

Como resultado desta pesquisa, esta seção apresenta a análise realizada sobre os 79 trabalhos completos relevantes, onde as questões de pesquisa são respondidas com base na literatura.

3.1 Análise da Revisão

A partir da leitura e extração dos dados relevantes dos artigos incluídos, foi possível representar a evolução das pesquisas relacionadas ao tema Sistemas de Informação Inteligentes no Brasil, e as principais instituições responsáveis por esta evolução.

A partir da observação do Gráfico 3, nota-se um grande aumento na quantidade de trabalhos voltados para o tema abordado no SBSI. Nas primeiras quatro edições estudadas, 2008 a 2011, foram identificados 21 trabalhos (27%) incluídos nesta revisão. Já entre os anos de 2012 e 2015 (nas últimas 4 edições do evento), este número quase triplicou (Gráfico 4), tendo sido identificados 58 trabalhos (73%), revelando assim um interesse crescente nas pesquisas no tema citado no Brasil.

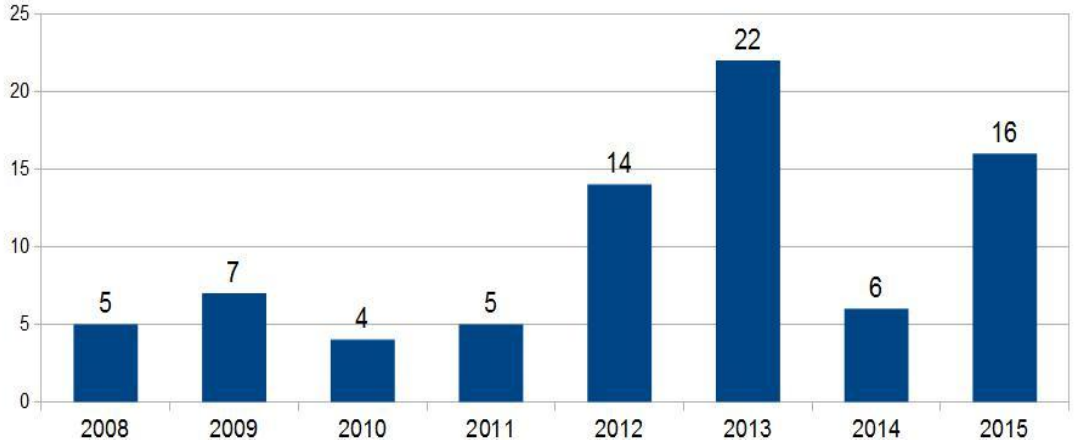


Gráfico 3. Artigos relevantes à pesquisa agrupados por ano de publicação.

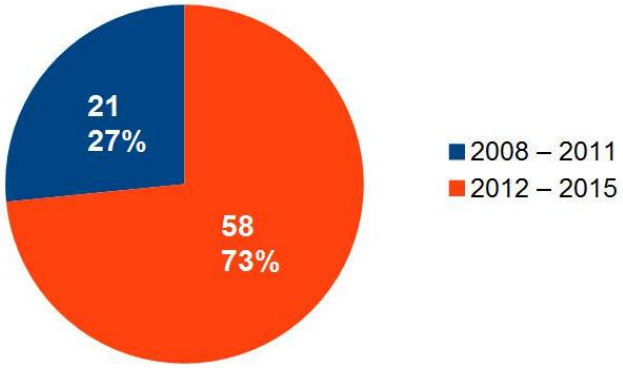


Gráfico 4. Artigos relevantes à pesquisa classificados por ano de publicação agrupados em dois quadriênios.

Em relação às instituições que desenvolveram os trabalhos utilizados nesta revisão, verificou-se que houve participações de 40 diferentes instituições, entre universidades federais, estaduais e institutos tecnológicos. Dado que há trabalhos feitos com parceria de autores de duas ou mais instituições, entre os anos de 2008 e 2011, houve trabalhos oriundos de 19 instituições. Este número subiu para 28 entre os anos de 2012 e 2015. O que representa um crescente interesse das instituições brasileiras no desenvolvimento de trabalhos e pesquisas no tema abordado neste trabalho.

Por outro lado, 41% desses trabalhos foram desenvolvidos por apenas três universidades brasileiras: USP, UFPE e UNIRIO. Essas instituições sustentam grande parte dos trabalhos incluídos nesta pesquisa (Gráfico 5) e são as principais responsáveis pela evolução nos últimos anos das pesquisas em Sistemas de Informação Inteligentes no Brasil. Isso pode ser explicado pelo fato de que essas universidades possuem reconhecimento nacional e internacional em pesquisas na área de Computação, em geral, e em Sistemas de Informação, em específico, uma vez que abrigam programas de pós-graduação com forte atuação em Sistemas de Informação.

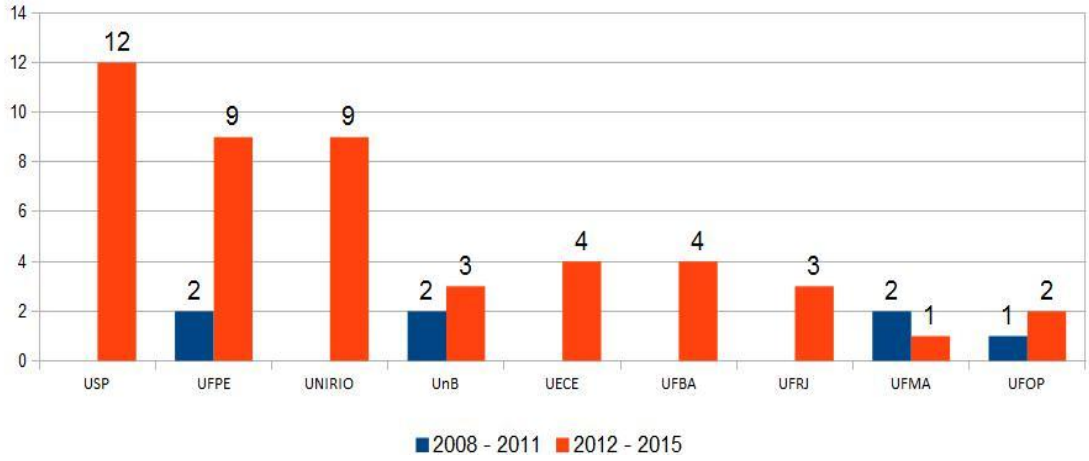


Gráfico 5. Universidades que mais publicaram artigos relevantes à pesquisa agrupados por dois quadriênios.

3.2 Questões de Pesquisa

De acordo com o estudo realizado, apresentamos nesta seção as questões de pesquisa definidas no protocolo e o mapeamento de dados em relação aos trabalhos revisados. Destacamos que todos os trabalhos relevantes abordaram todas as questões de pesquisa com excessão da questão de pesquisa 5 (QP5) do protocolo, que foi abordada de forma relevante à pesquisa em apenas 11% dos trabalhos.

QP1. Técnicas de Inteligência Artificial

A primeira questão de pesquisa refere-se a *Quais foram as técnicas de Inteligência Artificial utilizadas nos artigos*. Após a análise dos dados extraídos dos 79 artigos relevantes, foram identificadas seis principais técnicas que tiveram maior ocorrência nos trabalhos incluídos: mineração de dados, aprendizado de máquina, agentes inteligentes e sistemas multiagentes, algoritmos genéticos, redes neurais artificiais e ontologias, conforme ilustra o Gráfico 6.

Uma parcela desses artigos (23%) abordou duas ou mais técnicas. Dentre os 79 artigos utilizados nesta revisão, 26 deles (33%) abordaram mineração de dados para o reconhecimento de padrões, representando assim a técnica mais abordada nos trabalhos analisados nesta etapa. Este resultado pode ser explicado pela crescente preocupação apresentada nos artigos com o tratamento do grande volume de dados disponível na web ou em bancos de dados de organizações, que podem gerar informações de extrema importância para diversos usos como a tomada de decisão. Podem ser citados como exemplos os trabalhos de mineração de dados para predição, como desligamentos em sistemas elétricos [Maia et al. 2015], relacionamentos em redes sociais [Digiampietri et al. 2015], a ocorrência de moluscos em rios [Guandaline and Merschmann 2014], para o diagnóstico de doenças [Morais et al. 2012], como também a identificação de fatores que influenciam a evasão em cursos de graduação [Manhães et al. 2012].

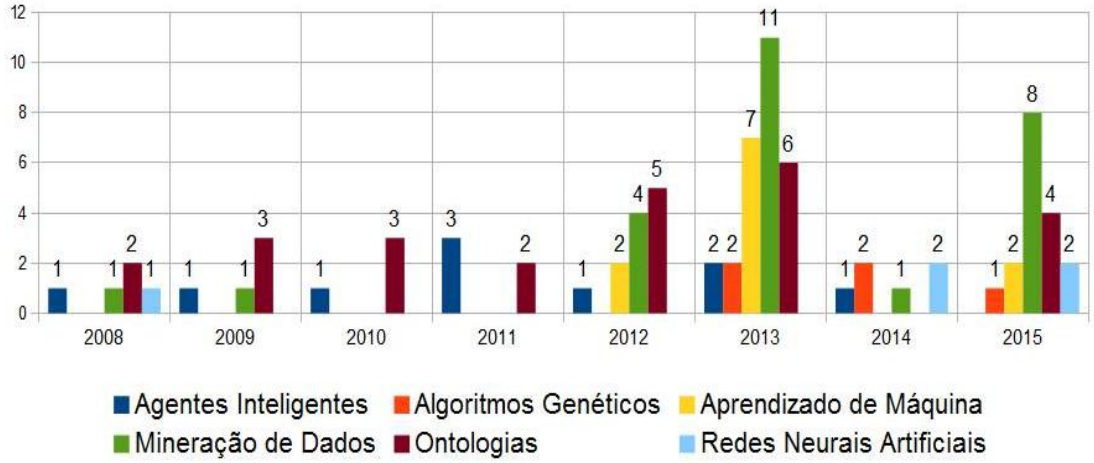


Gráfico 6. Técnicas de Inteligência Artificial abordadas pelos artigos relevantes à pesquisa agrupados por ano de publicação.

Presentes em 25 artigos (32%), as ontologias foram a segunda técnica mais abordada nos trabalhos, com representatividade em quase todas as edições do SBSI estudadas. Este fato pode ser explicado pela importante contribuição das ontologias na modelagem semântica e representação de conhecimento, fundamental para o desenvolvimento e implementação de diversas outras teorias e sistemas inteligentes como os voltados para o Reconhecimento de Padrões, Web Semântica e Processamento de Linguagem Natural. Podem ser citados os trabalhos que criam taxonomias para a classificação de estudos [Mota et al. 2013], ontologias aplicadas ao desenvolvimento de Sistemas de Informação Geográfica [Lamas et al. 2009], dentre outros. Mineração de dados e ontologias, juntos, abrangem 60% dos artigos relevantes nesta pesquisa.

Algoritmos genéticos e redes neurais artificiais, com 5 artigos cada uma, estão presentes em apenas 6% dos artigos, pois, para as tarefas de classificação ou otimização nas quais estes algoritmos foram utilizados nos trabalhos, alguns trabalhos optaram alternativamente pela escolha e avaliação de outros algoritmos como redes bayesianas, árvores de decisão, ou algoritmos de inteligência de enxame. Algoritmos genéticos foram aplicados em sistemas de gerenciamento de *workflow* [Malaquias et al. 2013], seleção de portfólio de investimentos [Tourinho et al. 2013], entre outros. Já as redes neurais foram aplicadas em trabalhos como sistemas de classificação de carcaças bovinas [Bittencourt et al. 2008] e sistemas de previsão do tempo de chegada de ônibus [Coquita et al. 2015].

QP2. Foco dos Trabalhos

A segunda questão de pesquisa trata de identificar se O foco do artigo está em representação de conhecimento, raciocínio automatizado ou ambos. Esta informação é representada no Gráfico 7.

Analisando o Gráfico 7 é possível perceber que a maioria (56%) dos artigos analisados nesta revisão tiveram como foco o raciocínio automatizado. O raciocínio automatizado consiste em usar informações armazenadas com a finalidade de responder a perguntas e tirar novas conclusões, e pode auxiliar nos processos de aprendizado de máquina, mineração de dados e também na comunicação entre agentes em um Sistema Multiagente como pode ser visto em [Dall'Oglio et al. 2010].

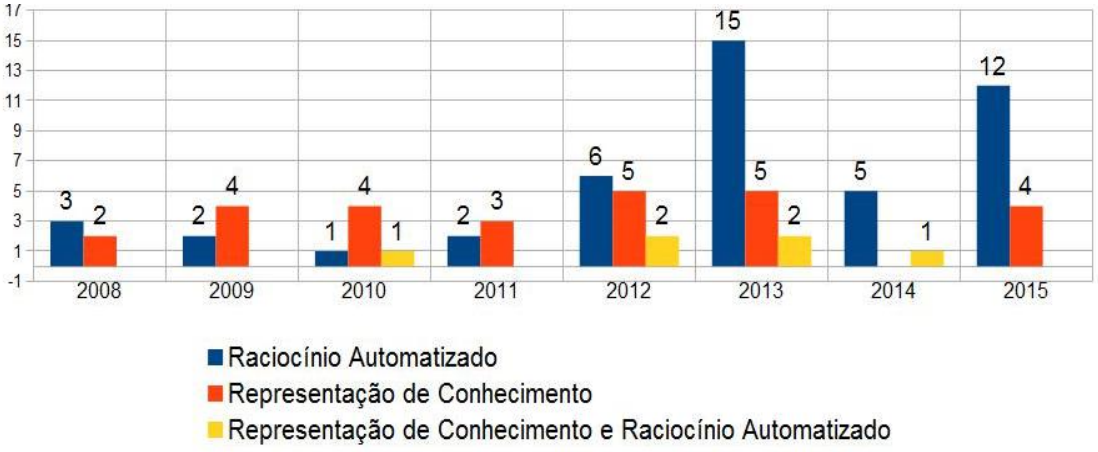


Gráfico 7. Foco dos artigos relevantes à pesquisa agrupados por ano.

A representação de conhecimento tem como objetivo o desenvolvimento de coleções estruturadas de informações e de conjuntos de regras de inferência que ajudem no processo de dedução automática para que seja administrado o raciocínio automatizado. Nos trabalhos analisados nesta revisão a representação de conhecimento foi abordada principalmente por meio de ontologias e foi o foco de 33% dos artigos, sendo utilizada para tarefas como modelagem semântica de sítios de conteúdo social [Menezes et al. 2010] e para dirimir ambiguidades na modelagem de processos de negócio [Martins et al. 2011], por exemplo.

Os artigos que abordaram ambos os focos em raciocínio automatizado e representação de conhecimento representam apenas 11% dos trabalhos utilizados nesta revisão, abordados nos trabalhos para atividades como o desenvolvimento de agentes inteligentes como agentes de apoio a gestão de projetos de software [Chagas et al. 2013] ou da mudança de requisitos [Dall'Oglio et al. 2010].

QP3. Principais Contribuições

A terceira questão de pesquisa é sobre *Quais são as principais contribuições do artigo*. De acordo com seu conteúdo, foram definidas 5 categorias de possíveis contribuições dos trabalhos:

- **Survey**: análise detalhada de uma área de pesquisa, com levantamento e comparação de trabalhos e identificação de questões em aberto;
- **Revisão**: revisão sistemática da literatura e classificação de trabalhos em subáreas;
- **Proposta**: proposição de uma abordagem em nível conceitual;
- **Modelo**: apresentação de um modelo e/ou arquitetura, sem implementação;
- **Implementação**: modelo e/ou arquitetura validada por implementação.

De acordo com o Gráfico 8, houve uma queda na quantidade de trabalhos do tipo Implementação na última edição do evento SBSI (2015), porém analisando os trabalhos das edições anteriores, é possível perceber que os trabalhos do tipo Implementação como [Coquita et al. 2015], [Blomberg and Ruiz 2013], [Tourinho et al. 2013], [Frizzarini and Lauretto 2013] entre outros, tiveram um grande aumento, chegando a ultrapassar em algumas edições a quantidade de trabalhos do tipo Modelo como os artigos de [Santos and Girardi 2012], [Rigo et al. 2008] e [Lamas et al. 2009] por exemplo, que representam o tipo predominante com 29 trabalhos (37%). Os

trabalhos de Implementação são representados por 27 trabalhos (34%), fato este que sugere um processo de amadurecimento nos trabalhos em Sistemas de Informação Inteligentes no contexto brasileiro.

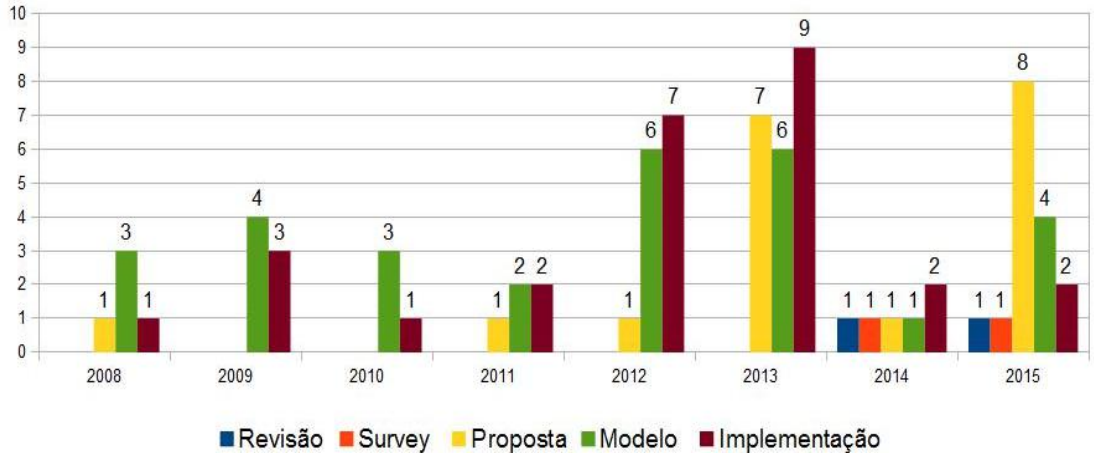


Gráfico 8. Contribuições dos artigos relevantes à pesquisa agrupados por ano.

Outro fato que pode ser observado a partir da observação do gráfico é o surgimento de trabalhos do tipo Revisão Sistemática como em [Trucolo and Digiampietri 2014], e *Survey*, como em [Guandaline and Merschmann 2014], referentes ao tópico-alvo desta pesquisa nos últimos anos, o que pode sugerir que há uma crescente abordagem sobre Sistemas de Informação Inteligentes na literatura, que serve como base para o desenvolvimento destes tipos de trabalho.

QP4. Domínios de Aplicação

A quarta questão de pesquisa refere-se a *Qual é o domínio de aplicação do trabalho descrito no artigo.* Os trabalhos analisados abordaram uma quantidade de domínios bastante diversificado, porém os principais domínios de aplicação identificados nos trabalhos incluídos nesta revisão estão representados no Gráfico 9.

- **Gestão do Conhecimento:** principalmente a gestão de informação organizacional [Confort et al. 2015], para a criação e utilização de conhecimento para apoio em tomadas de decisão por meio do desenvolvimento de BI's (*Business Intelligence*) [Pintas et al. 2011], entre outros.
- **Web Semântica:** alguns trabalhos apresentaram ferramentas e técnicas para atividades como a composição de serviços Web semânticos [Fonseca et al. 2009], desenvolvimento de sistemas de transporte coletivo inteligentes [Veiga and Bulcão-Neto 2015] e para o desenvolvimento de arquiteturas para aplicações baseadas em Web Semântica [Rigo et al. 2008].
- **Educação:** a educação foi outro domínio bastante abordado nos trabalhos, principalmente voltada para o ensino superior, através de ferramentas de apoio à identificação de fatores que influenciam a evasão de cursos [Manhães et al. 2012], ferramentas de apoio à educação a distância [Penedo and Capra 2012] e ao acompanhamento de projetos científicos, entre outras.
- **Saúde:** muitas técnicas e ferramentas têm sido propostas neste domínio, como aquelas aplicadas a cuidados com a saúde por meio de dispositivos móveis (*m-Health*) e ao diagnóstico de doenças, como a asma [Morais et al. 2012].

- **Processos de Negócios:** há esforços direcionados para processos de negócios organizacionais na forma de propostas para a formação e eliminação de ambiguidades em processos de negócios [Martins et al. 2011] ou para o apoio a mudanças nesses processos.
- **Processamento de Imagem:** há trabalhos voltados para o desenvolvimento de sistemas para processamento de imagens, como aquelas aplicadas em atividades de reconhecimento de LIBRAS (Língua Brasileira de Sinais) [Digiampietri et al. 2012] e de classificação de carcaças bovinas [Bittencourt et al. 2008].

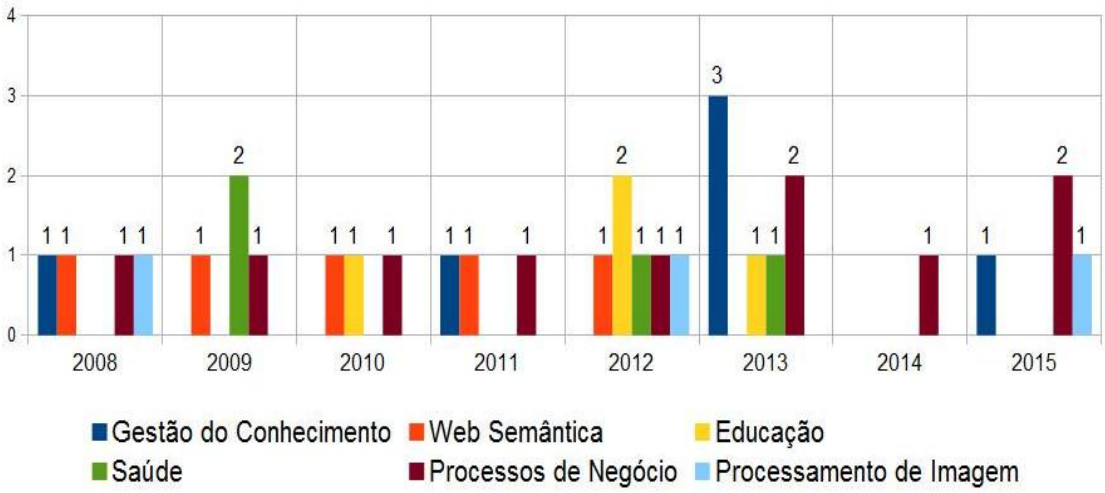


Gráfico 9. Domínios de aplicação dos artigos relevantes à pesquisa agrupados por ano.

QP5. Questões em Aberto

A última questão de pesquisa trata *Quais são as principais questões em aberto descritas nos trabalhos.* As questões em aberto são aquelas que não têm sido atacadas pela literatura, mas que são destacadas pelos autores como áreas que podem trazer contribuições promissoras. As principais questões em aberto encontradas nos trabalhos relacionados a área de Sistemas de Informação Inteligentes são descritas a seguir e também resumidas na Tabela 1.

- **Meta-heurísticas para cenários dinâmicos:** existem propostas de algoritmos de otimização voltados para problemas dinâmicos, em cenários que estão em constante mudança. Porém, ainda existe a necessidade de algoritmos mais eficientes ou a implementação e o aperfeiçoamento de algoritmos diferentes dos algoritmos gulosos, como os de inteligência de enxame, que possam apresentar melhores desempenhos para estes problemas. Isto pode ser evidenciado nos trabalhos de [Barbosa et al. 2015] e [Lima et al. 2012], para prover sistemas de roteirização mais eficazes e eficientes.
- **Tratamento de dados faltantes em algoritmos de classificação:** como se pode observar em [Guandaline and Merschmann 2014], [Frizzarini and Lauretto 2013] e [Blomberg and Ruiz 2013], existe uma grande necessidade de se criar ferramentas e métodos para o tratamento ou substituição de dados faltantes ou desbalanceados para sua utilização nas tarefas de mineração de dados, mais especificamente para o uso em árvores de classificação.
- **Replicação de experimentos com dados dinâmicos:** segundo [Trucolo and Digiampietri 2014], há uma grande dificuldade na replicação de experimentos

que utilizam dados dinâmicos, como experimentos de sistemas que utilizam dados extraídos de redes sociais, pois nem sempre é possível obter os mesmos dados para replicá-los e realizar a avaliação destes.

- **Funções para alinhamento de ontologias:** de acordo com [Alves et al. 2013] e [Confort et al. 2015], há a necessidade de se criar funções genéricas para avaliação e classificação de alinhamento de ontologias para tratar a heterogeneidade semantica. A literatura classifica as funções existentes como de baixa precisão.
- **Métodos de recuperação de ontologias:** há a necessidade de desenvolver métodos de recuperação de ontologias de aplicação para a realização de buscas semânticas em repositórios de ontologias para a extensão de ontologias de aplicação desenvolvidas [Santos and Girardi 2012].

Tabela 1. Trabalhos relevantes à pesquisa agrupados pelas principais questões de pesquisa em aberto encontradas.

Questões de Pesquisa	Trabalho(s)
Meta-heurísticas dinâmicas	[Barbosa et al. 2015], [Lima et al. 2012]
Dados faltantes em algoritmos	[Guandaline and Merschmann 2014], [Frizzarini and Lauretto 2013], [Blomberg and Ruiz 2013]
Replicação de experimentos com dados dinâmicos	[Trucolo and Digiampietri 2014]
Funções para alinhamento de ontologias	[Alves et al. 2013], [Confort et al. 2015]
Métodos de recuperação de ontologias	[Santos and Girardi 2012]

4. Considerações Finais

A proposta deste trabalho foi apresentar um processo de revisão sistemática da literatura para levantar dados e analisar o estado da arte das pesquisas em Sistemas de Informação Inteligentes no Brasil, respondendo às seguintes questões de pesquisa:

1. Quais são as teorias de Inteligência Artificial utilizadas nos artigos?
2. O foco dos artigos está em representação de conhecimento, raciocínio automatizado, ou ambos?
3. Quais são as principais contribuições dos artigos: *survey*, revisão sistemática, proposta conceitual, modelo ou implementação?
4. Quais são os domínios de aplicação dos trabalhos?
5. Quais são as principais questões de pesquisa em aberto descritas nos trabalhos?

A revisão sistemática considerou os trabalhos publicados nas edições de 2008 a 2015 do Simpósio Brasileiro de Sistemas de Informação, relacionados com o tema de Sistemas Inteligentes e presentes na biblioteca digital BDBComp. Cada artigo retornado da base teve sua relevância analisada de acordo com os critérios de exclusão; os não relevantes foram eliminados. O restante seguiu para as demais fases do processo de revisão e serviu de base para o levantamento dos dados necessários para responderem às questões de pesquisa e, assim, revelar o estado da arte das pesquisas no tema referido.

Dos 106 artigos retornados inicialmente, um total de 79 (69%) foi considerado relevante para este trabalho. A partir destes artigos foi possível perceber que:

- a curva de crescimento de publicações sobre Sistemas Inteligentes no Brasil revela um interesse cada vez maior nesta área;
- o número de instituições que desenvolveram trabalhos no tema abordado mais que dobrou nos últimos 4 anos, porém um número reduzido de universidades brasileiras sustenta a maioria dessas pesquisas;
- mineração de dados (Reconhecimento de Padrões) e ontologias (Representação de Conhecimento) são as teorias de Inteligência Artificial mais abordadas nos artigos analisados;
- o foco das pesquisas em Sistemas de Informação Inteligentes no Brasil está mais voltado para o raciocínio automatizado do que para a representação de conhecimento;
- a curva de crescimento dos trabalhos do tipo Implementação (I) revela que as pesquisas em Sistemas de Informação Inteligentes no Brasil estão passando por um processo de amadurecimento;
- gestão do conhecimento, processos de negócios e saúde são os domínios de aplicação mais abordados pelos trabalhos, sendo processos de negócios o domínio mais pesquisado;
- as principais questões em aberto encontradas nos trabalhos estão voltadas para o desenvolvimento e a aplicação de meta-heurísticas mais eficientes em cenários dinâmicos, o tratamento de dados faltantes ou desbalanceados em algoritmos de classificação, aos problemas de replicação de experimentos com dados dinâmicos e à necessidade de funções mais eficazes para o alinhamento de ontologias.

A síntese do estado da arte das pesquisas em Sistemas de Informação Inteligentes obtida como resultado deste trabalho tem como objetivo subsidiar novas pesquisas na área citada no Instituto de Informática da Universidade Federal de Goiás. A lista completa dos artigos relevantes utilizados nesta pesquisa, classificados por teoria, foco, categoria e domínio de aplicação, encontra-se em: docs.google.com/document/d/1lXwo3b5h8sYnTsucnXjl9yCj5INR0rbIzgm_Wh70C7A.

Agradecimentos

Aos coordenadores das edições do Simpósio Brasileiro de Sistemas de Informação: Fátima de Lourdes Santos Nunes, Fernanda Araújo Baião, Flávia Maria Santoro, Renata Mendes de Araújo, Rita Suzana Pitangueira Maciel, Sean Wolfgand Matsui Siqueira, Valdemar Vicente Graciano Neto, Vaninha Vieira dos Santos e Vinícius Sebba Patto.

Referências

ALVES, A. et al. (2013). Classificador de alinhamento de ontologias utilizando técnicas de aprendizado de máquina. In: Anais do IX Simpósio Brasileiro de Sistemas de Informação. João Pessoa-PB, Brasil. p. 25-36.

BARBOSA, D. et al. (2015). Aplicação da otimização por colônia de formigas ao problema de múltiplos caixeiros viajantes no atendimento de ordens de serviço nas empresas de distribuição de energia elétrica. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Goiânia-GO, Brasil. p. 23-30.

BITTENCOURT, C. et al. (2008). Sistema de Classificação Automática de Carcaças Bovinas. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Rio de Janeiro-RJ, Brasil. p. 1-10.

BLOMBERG, L. & RUIZ, D. (2013). Evaluating the Influence of Missing Data on Classification Algorithms in Data Mining Applications. In: Anais do IX Simpósio Brasileiro de Sistemas de Informação. João Pessoa-PB, Brasil. p. 734-743.

CHAGAS, J. et al. (2013). Uma abordagem Baseada em Agentes de Apoio à Gestão de Projetos de Software. In: Anais do IX Simpósio Brasileiro de Sistemas de Informação. João Pessoa-PB, Brasil. p. 625-636.

CONFORT, V. et al. (2015). Extração de Ontologias a partir de Histórias: um estudo exploratório em storytelling. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Goiânia-GO, Brasil. p. 243-250.

COQUITA, K. et al. (2015). Sistema de Previsão do Tempo de Chegada dos Ônibus Baseado em Dados Históricos Utilizando Modelos de Regressão. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Goiânia-GO, Brasil. p. 119-126.

DALL'OGLIO, P. et al. (2010). ChangeMan: Um Sistema Multi-Agente para Gestão da Mudança de Requisitos com suporte à Rastreabilidade e Análise de Impacto. In: Anais do VI Simpósio Brasileiro de Sistemas de Informação. Marabá-PA, Brasil. p. 1-12.

DIGIAMPIETRI, L. et al. (2015). Um Sistema de Predição de Relacionamentos em Redes Sociais. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Goiânia-GO, Brasil. p. 139-146.

FONSECA, A. et al. (2009). Gerenciando o desenvolvimento de uma composição de Serviços Web Semânticos através do OWL-S Composer. In: Anais do V Simpósio Brasileiro de Sistemas de Informação. Brasília-DF, Brasil. p. 109-120.

FRIZZARINI, C.; LAURETTO, M. (2013). Proposta de um Algoritmo para Indução de Árvores de Classificação para Dados Desbalanceados. In: Anais do IX Simpósio Brasileiro de Sistemas de Informação. João Pessoa-PB, Brasil. p. 722-733.

GUANDALINE, V.; MERSCHMANN, L. (2014). Avaliação de técnicas de mineração de dados para predição de ocorrência do mexilhão dourado em rios da América Latina. In: Anais do X Simpósio Brasileiro de Sistemas de Informação. Londrina-PR, Brasil. p. 663-674.

KITCHENHAM, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(2004):1–26.

LAMAS, A. et al. (2009). Ontologias e Web Services aplicados ao desenvolvimento de Sistemas de Informação Geográfica Móveis Sensíveis ao Contexto. In: Anais do V Simpósio Brasileiro de Sistemas de Informação. Brasília-DF, Brasil. p. 157-168.

LIMA, V. et al. (2012). UbibusRoute: Um Sistema de Identificação e Sugestão de Rotas de Ônibus Baseado em Informações de Redes Sociais. In: Anais do VIII Simpósio Brasileiro de Sistemas de Informação. São Paulo-SP, Brasil. p. 516-527.

MAIA, A. et al. (2015). Avaliação de Técnicas de Mineração de Dados para Predição de Desligamentos em Sistemas Elétricos de Potência. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Goiânia-GO, Brasil. p. 155-162.

MALAQUIAS, F. et al. (2013). Sistema de Gerenciamento de Workflow baseado em Redes de Petri e em Algoritmos Genéticos. In: Anais do IX Simpósio Brasileiro de Sistemas de Informação. João Pessoa-PB, Brasil. p. 415-426.

MANHÃES, L. et al. (2012). Identificação dos Fatores que Influenciam a Evasão em Cursos de Graduação Através de Sistemas Baseados em Mineração de Dados: Uma Abordagem Quantitativa. In: Anais do VIII Simpósio Brasileiro de Sistemas de Informação. São Paulo-SP, Brasil. p. 468-479.

MARTINS, A. et al. (2011). Uso de uma Ontologia de Fundamentação para Dirimir Ambiguidades na Modelagem de Processos de Negócio. In: Anais do VII Simpósio Brasileiro de Sistemas de Informação. Salvador-BA, Brasil. p. 129-140.

MENEZES, L. et al. (2010). Uma Abordagem Baseada em Ontologias para Modelagem Semântica e Geográfica de Sites de Conteúdo Social no Contexto dos Setores Públicos. In: Anais do VI Simpósio Brasileiro de Sistemas de Informação. Marabá-PA, Brasil. p. 1-12.

MORAIS, D. et al. (2012). Sistema Móvel de Apoio a Decisão Médica Aplicado ao Diagnóstico de Asma - InteliMED. In: Anais do VIII Simpósio Brasileiro de Sistemas de Informação. São Paulo-SP, Brasil. p. 528-539.

MOTA, M. et al. (2013). Uma Taxonomia Ontológica para a Classificação de Estudos em Informática Médica. In: Anais do IX Simpósio Brasileiro de Sistemas de Informação. João Pessoa-PB, Brasil. p. 589-600.

PENEDO, J.; CAPRA, E. (2012). Mineração de Dados na Descoberta do Padrão de Usuários de um Sistema de Educação à Distância. In: Anais do VIII Simpósio Brasileiro de Sistemas de Informação. São Paulo-SP, Brasil. p. 396-407.

PINTAS, J. et al. (2011). O Papel da Semântica no Business Intelligence 2.0: Um Exemplo no Contexto de um Programa de Pós-Graduação. In: Anais do VII Simpósio Brasileiro de Sistemas de Informação. Salvador-BA, Brasil. p. 69-80.

RIGO, S. et al. (2008). Arquitetura baseada em Web Semântica para aplicações de Hipermídia Adaptativa. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Rio de Janeiro-RJ, Brasil. p. 1-10.

RUSSELL, S.J.; NORVIG, P. (2003). Artificial Intelligence: A Modern Approach, 2nd. Edition, Prentice Hall.

SAMPAIO, R. F.; ANCINI, M. C. (2007). Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. Revista Brasileira de Fisioterapia, vol. 11, pp. 83–89.

SANTOS, L.; GIRARDI, R. (2012). Uma Técnica Orientada por Objetivos para a Construção de Ontologias de Aplicação. In: Anais do VIII Simpósio Brasileiro de Sistemas de Informação. São Paulo-SP, Brasil. p. 13-24.

SCHUTZER, D. (1987). Artificial intelligence: an applications-oriented approach. New York: Van Nostrand Reinhold Company.

SIRIN Evren, et al. (2007). "Pellet: A Pratical OWL-DL Reasoner", Valencia, Journal of Web Semantics. p. 1-26.

TRUCOLO, C.; DIGIAMPIETRI, L. (2014). Uma Revisão Sistemática acerca das Técnicas de Identificação e Análise de Tendências. In: Anais do X Simpósio Brasileiro de Sistemas de Informação. Londrina-PR, Brasil. p. 639-650.

TOURINHO, R. et al. (2013). Um Algoritmo Genético para Seleção de Portfólio de Investimentos com Restrições de Cardinalidade e Lotes-Padrão. In: Anais do IX Simpósio Brasileiro de Sistemas de Informação. João Pessoa-PB, Brasil. p. 541-552.

VEIGA, E. F.; BULCÃO NETO, R. F. (2015). Uma abordagem ontológica aplicada ao gerenciamento de informação de redes de transporte coletivo. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. Goiânia-GO, Brasil. p. 485-492.

Um serviço de representação ontológica de informações de localização para aplicações em Internet das Coisas

José Augusto Batista Neto, Ernesto Fonseca Veiga, Renato de Freitas Bulcão Neto[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.690-900 – Goiânia – GO – Brazil

{josebatista,ernestofonseca,renato}@inf.ufg.br

Abstract. *In the Internet of Things paradigm, where information is produced by several types of sensors and disseminated via the web, interoperability becomes an essential requirement. For location information, which are used for applications in different fields, have common understanding of its meaning, this paper proposes a semantic interoperable representation service of location information, based on consolidated ontologies and recommended by the literature. The proposal was validated through a case study in location of cities. The main contributions of this work are: a reusable service for representing location information in interoperable format and expressive; and supporting the development of applications using location information, focused on the IoT paradigm.*

Resumo. *No paradigma de Internet das Coisas, onde informações são produzidas pelos mais diversos tipos de sensores e disseminadas via web, a interoperabilidade torna-se um requisito essencial. Para que informações de localização, que são utilizadas por aplicações de diferentes domínios, possuam interpretação uniforme de seu significado, este trabalho propõe um serviço de representação semântica e interoperável de informações de localização, fundamentado em ontologias consolidadas e recomendadas pela literatura. A proposta foi validada através de um estudo de caso em localização de cidades. As principais contribuições deste trabalho são: um serviço reutilizável para representação de informações de localização em formato interoperável e expressivo; e o apoio ao desenvolvimento de aplicações que utilizam informações de localização, voltadas para o paradigma de IoT.*

1. Introdução

O Dr. Mário Amato é diretor geral de uma central de monitoramento remoto de pacientes, cujo trabalho é acompanhar em tempo real os sinais vitais dos pacientes que se encontram em suas residências. A central de monitoramento é dividida em 3 bases, localizadas em pontos estratégicos da cidade, cada qual com um conjunto de médicos e enfermeiros. Quando necessário, um conjunto de profissionais precisa se deslocar até os pacientes para atender suas necessidades, seja quanto a procedimentos de rotina ou de caráter emergencial, como por exemplo, a aferição de anormalidades em sinais vitais que geram alterações de estado clínico.

Para otimizar o processo de atendimento aos pacientes das diferentes partes da cidade, foi implantado um ambiente computacional, voltado para Internet das Coisas (IoT), onde todos os membros da equipe médica utilizam em seus *smartphones*, um aplicativo

denominado Sistema para Gerenciamento de Atendimentos e Emergências (SiGAE). Este aplicativo troca informações com o Sistema de Monitoramento de Pacientes (SiMP), implantado na residência e nos dispositivos de cada paciente, que acompanha tanto as suas informações clínicas quanto a sua localização na cidade.

Desse modo, quando o SiMP detecta que um determinado paciente precisa de atendimento, ele busca as informações de localização dos profissionais da equipe, acionando um alarme para os membros da base mais próxima do paciente em questão, através do SiGAE, fornecendo ainda a localização e as informações clínicas sobre o estado atual do paciente. Uma vez que o SiMP tem acesso à localização tanto dos profissionais de saúde quanto dos pacientes, ele pode gerenciar, de maneira inteligente e automática, para que uma equipe ou um serviço de emergência chegue o mais rápido possível até o local do atendimento, distribuindo os membros da equipe para diferentes procedimentos, inibindo conflitos e agilizando o processo.

Este cenário apresenta claramente um exemplo da troca de informações entre diferentes aplicações, destacando a interoperabilidade entre os sistemas SiGAE e SiMP. Para que, nessa troca de informações, todos os dados compartilhem da mesma compreensão, independente do sistema ou aplicação, é necessário um mecanismo de representação uniforme, que garanta uma interpretação homogênea. Para cumprir este requisito, a literatura em IoT têm destacado e recomendado a representação baseada em ontologias [Perera et al. 2014].

Ontologia é uma metodologia adotada para modelagem de informações, agregando semântica para que, junto com seus metadados, possuam significado de maneira não ambígua, de acordo com o contexto ao qual elas pertencem. Ontologias descrevem domínios específicos de maneira formal e abstrata, garantindo que as informações tenham um sentido único através de definições de classes, indivíduos, atributos, axiomas e relacionamentos, permitindo seu reuso, desacoplamento e interoperabilidade [Bastos et al. 2013]

No intuito de facilitar o compartilhamento e o gerenciamento dessas informações, também denominadas contexto, em ambientes de IoT, estas informações são enviadas para Sistemas de Gerenciamento de Contexto (CMS), que têm como objetivo principal oferecer serviços como representação das informações, comunicação, histórico, entre outros, contribuindo para o desenvolvimento de aplicações. *Hermes* é um CMS baseado em ontologias, proposto por [Veiga et al. 2014], que oferece serviços de gerenciamento das informações de contexto para ambientes de IoT, entre eles a representação ontológica, garantindo interoperabilide e expressividade dessas informações.

O componente de *Hermes* que realiza a representação do contexto, chamado *Hermes Widget* (HW), é essencial para que os demais serviços de *Hermes* e aplicações possam utilizar o contexto de maneira interoperável, recebendo as informações de sensores do ambiente e representando-as com base em ontologias. [Veiga et al. 2016] apresentam um estudo de caso em monitoramento de sinais vitais, onde foram criados HWs para representação ontológica de diferentes informações, como: pressão arterial, temperatura corpórea, frequência de pulso, entre outras. No entanto, não se sabe de qual lugar no mundo essas informações estão sendo enviadas, por exemplo, se estão vindo de sensores instalados em uma sala/prédio, ou de sensores móveis, como em um celular ou veículo.

A informação de localização possui importância primordial em diferentes domínios de aplicação, como no cenário apresentado, onde, além de informações sobre o estado clínico de um determinado paciente, a equipe médica precisa saber exatamente onde ele está localizado na cidade, para que o atendimento possa ser realizado em tempo hábil. Tendo em vista a necessidade de informações contextuais de localização, o objetivo deste trabalho é o desenvolvimento de um serviço de representação ontológica para informações de localização. Este serviço, nomeado *HW Locator*, tem como base o componente *Hermes Widget*, e utiliza ontologias amplamente adotadas na literatura.

O artigo está organizado da seguinte forma: A Seção 2 apresenta *Hermes Widget* e as ontologias utilizadas na proposta; a Seção 3 discorre sobre os requisitos, arquitetura e implementação do *HW Locator*, além de um estudo de caso para sua utilização; a Seção 4 realiza uma comparação com trabalhos relacionados; e a Seção 5 conclui o trabalho.

2. Fundamentação Teórica

Com o objetivo de fornecer as informações necessárias para a compreensão do trabalho, essa seção apresenta o *Hermes Widget* e as ontologias SSN, DUL e WGS84.

2.1. *Hermes Widget*

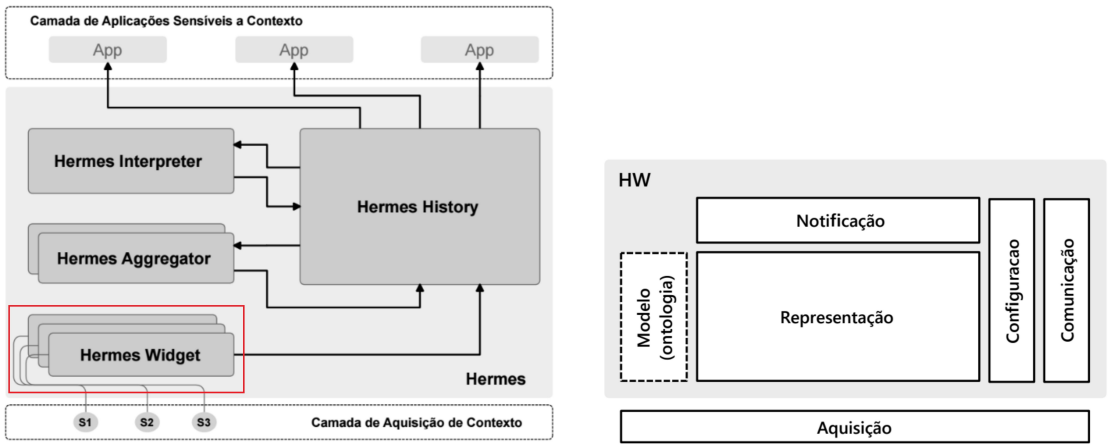
Hermes é um CMS baseado em ontologias, proposto por [Veiga et al. 2014], que controla e orquestra serviços relacionados às etapas do ciclo de vida das informações de contexto destacadas por [Perera et al. 2014]: aquisição, modelagem, interpretação e disseminação. A arquitetura de *Hermes* é apresentada na Figura 1(a), e os seus principais requisitos são: apoiar gerenciamento de informações de contexto; arquitetura orientada a componentes; e fornecimento de serviços baseados na semântica das informações, tais como representação, inferência, filtragem e histórico [Veiga et al. 2014].

Hermes Widget (HW), destacado na Figura 1(a), é o componente responsável pelo serviço de representação de contexto no CMS *Hermes*. Essencial para atribuir interoperabilidade às informações do ambiente, ele recebe dados de leitura de sensores como entrada, e realiza sua representação com base nas ontologias do domínio. Dessa forma, uma informação que possui inicialmente baixa expressividade e diferentes possibilidades de interpretação, ganha semântica e pode ser utilizada por diferentes aplicações, de maneira uniforme [Veiga and Bulcão Neto 2016].

A arquitetura do HW, apresentada na Figura 1(b), mostra a ligação entre a camada de representação e os modelos ontológicos, pré-definidos de acordo com o domínio de aplicação. As representações realizadas por este serviço são gerenciadas pelas camadas de *configuração*, *comunicação* e *notificação*, sendo as duas últimas responsáveis por entregar o contexto representado aos demais componentes de *Hermes* (para interpretação, histórico, etc.) e às aplicações.

Para que aplicações e componentes interessados possam receber as informações de contexto representadas por HW, este componente possui um arquivo de configuração, que está definido na camada de nome análogo, e contém os parâmetros para assinatura de tópicos criados para as instâncias do componente, no formato *JSON* [Veiga and Bulcão Neto 2016].

Em ambientes de IoT, uma grande quantidade de informações são produzidas por uma gama de sensores distintos, embarcados em dispositivos heterogêneos. Uma mesma



(a) (b)

Figura 1. (a) Arquitetura do CMS *Hermes*; e (b) em destaque, a arquitetura do componente *Hermes Widget* [Veiga and Bulcão Neto 2016] .

informação muitas vezes precisa ser consumida por diferentes aplicações, até mesmo de diferentes domínios. Por ser representada com base em ontologias, a informação produzida por HWs é altamente expressiva e interoperável, possuindo o mesmo significado independente da aplicação.

2.2. Ontologia SSN (*Semantic Sensor Network*)

Com objetivo de promover a integração das informações obtidas de diferentes sensores, e também a interoperabilidade entre sistemas e aplicações, o W3C *Incubator Group* propõe a ontologia *Semantic Sensor Network* (SSN). A SSN possui construtores que permitem descrever observações (aferições), precisão, capacidade, e métodos que os sensores utilizam para detectar o ambiente em que estão inseridos. Como elemento central para representação semântica de dados de sensoriamento, a SSN utiliza o padrão Estímulo-Sensor-Observação (SSO), apresentado na Figura 2 [Compton et al. 2012].

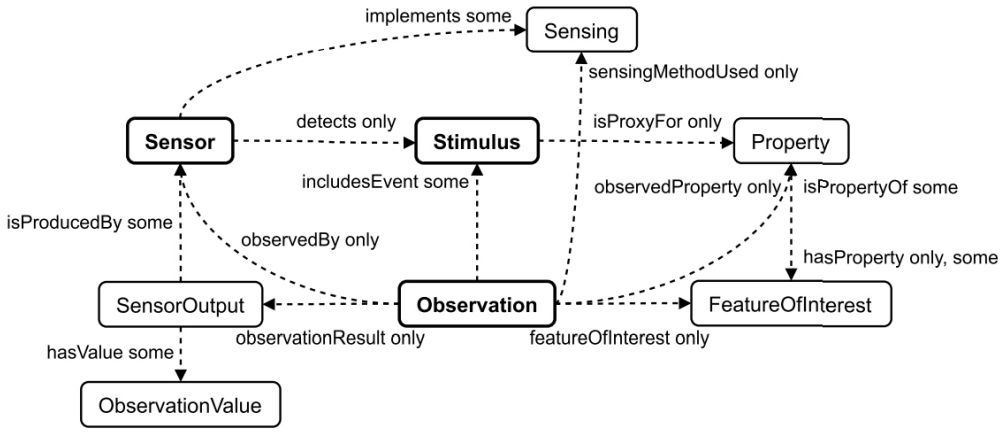


Figura 2. O Padrão Estímulo-Sensor-Observação [Compton et al. 2012]

Com o uso do padrão SSO, a ontologia SSN pode ser vista por quatro diferentes perspectivas: *i)* do sensor, com foco sobre o que é sensoreado, como é sensore-

ado e o que é detectado por ele; *ii)* da observação, com foco em aferições de dados e os metadados correspondentes; *iii)* do sistema, com foco em sistemas de sensores e sua implantação; e *iv)* das características e propriedades de interesse, que o sensor captura [Compton et al. 2012].

Dessa forma, o padrão SSO é o ponto central para a representação e realização das ligações entre o sensor, a propriedade sensoreada, a entidade de interesse, e o dado aferido, abrangendo as perspectivas acima descritas [Compton et al. 2012].

2.3. Ontologia DUL (DOLCE + DnS Ultralite)

Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) é uma ontologia de nível superior desenvolvida pelo Laboratório de Ontologia Aplicada *(LOA)*. Ela faz a representação ontológica da linguagem natural subjacente e de senso comum humano, por exemplo, a classe *Physical Object*[1] que define o que podem ser objetos físicos no mundo através de axiomas. A DOLCE se propõe ser um ponto de partida para promover ligações ou relações entre outras ontologias [Gangemi et al. 2002].

DOLCE+DnS Ultralite (DUL)[2] é uma ontologia criada a partir da junção das ontologias DOLCE e DnS Utralite, que fornece um conjunto de conceitos de nível superior, facilitando a descrição de informações e a interoperabilidade entre várias ontologias de nível inferior. Através de suas classes e propriedades, realiza a ligação entre ontologias, como exemplo, a ligação entre uma ontologia que realiza representação de sensores e uma ontologia para informações de localização, permitindo descrever em que local do espaço foi realizada uma aferição de um determinado sensor.

A Figura 3 apresenta as principais classes da ontologia DUL e o seu alinhamento com a ontologia SSN. A SSN foi criada com base em ontologias de referência, e possui diversas classes que são especializações de classes da ontologia DUL.

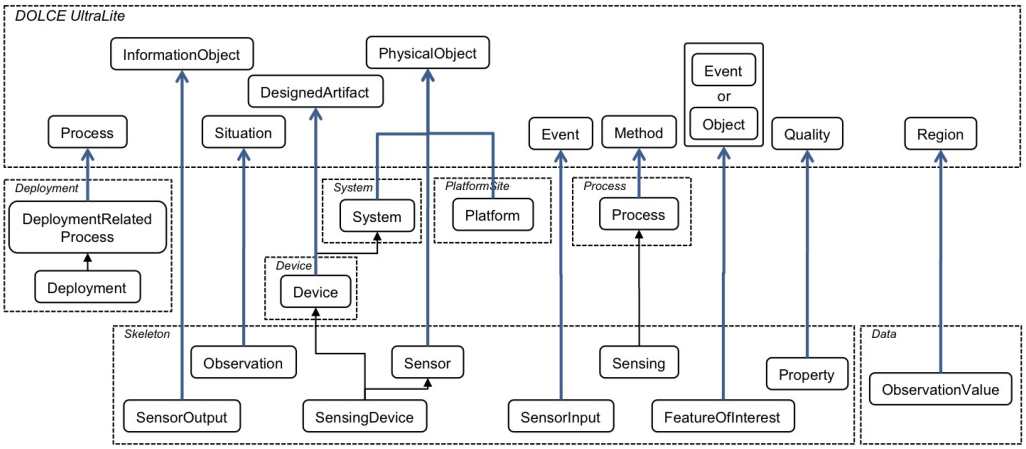


Figura 3. Alinhamento entre as ontologias DUL e SSN [Compton et al. 2012]

2.4. Ontologia WGS84

Criado em 1984 o *World Geodetic System (WGS)* é um padrão utilizado para fazer referência às coordenadas da Terra, cartografia, navegação incluindo também o *Global Po-*

[1]http://www.loa.istc.cnr.it/ontologies/DOLCE-Lite.owl#physical-object
[2]http://ontologydesignpatterns.org/wiki/Ontology:DOLCE%2BDnS_Ultralite

sitioning System (GPS). O padrão define valores de latitude, longitude e altitude como principais parâmetros para se definir um ponto. Sua origem é a massa central da Terra e os eixos Z, X e Y são, respectivamente, seus pontos de partida: *i)* polo norte; *ii)* linha do equador e seu ponto de partida o meridiano de greenwich; e *iii)*, a 90º leste do eixo X na linha do equador [Kumar 1988].

Para que sistemas possam utilizar essas informações de localização de maneira interoperável, foi criada pela W3C a ontologia WGS84[3], que representa *Pontos* contendo *latitude*, *longitude*, *altitude* e outras informações sobre objetos/coisas espacialmente localizados. Essa ontologia permite não somente representar mapas, mas também entidades que estão em um mapa, utilizando outras ontologias para realizar ligações.

A Figura 4 apresenta as principais classes da ontologia WGS84, bem como a descrição do indivíduo *coordenada*, que é uma instância da classe *Point*, que possui as propriedades latitude e longitude.

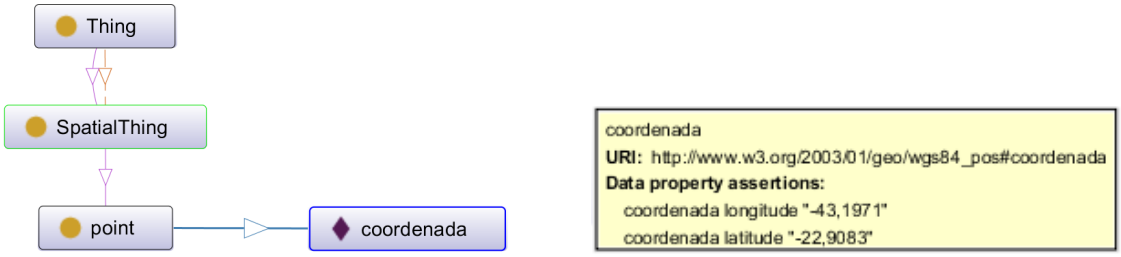


Figura 4. Principais classes da WGS84 e descrição do indivíduo *coordenada*.

Várias aplicações amplamente utilizadas, como por exemplo, o *Google Maps*[4], que possui um sistema de coordenadas para referenciar exclusivamente um ponto no mundo, utilizam como padrão o WGS84. O aplicativo *Instagram*[5] também utiliza o WGS84 para adicionar um local a uma mídia, fotos ou vídeos, e depois através das coordenadas de latitude e longitude conseguem encontrar locais e mídias recentes em determinadas áreas.

3. Representação semântica e interoperável de informações de localização

Esta seção apresenta o trabalho desenvolvido, abrangendo desde os requisitos, até a implementação e validação do serviço de representação ontológica de localização.

3.1. Requisitos e Projeto Arquitetural

Por ser baseado no componente *Hermes Widget*, o *HW Locator* herda o seus requisitos, que são [Veiga et al. 2016]:

- Permitir a especificação de um sensor e do processo de sensoreamento: o componente deve permitir a descrição de um processo de sensoreamento e de suas relações, tratando termos e eventos deste domínio tais como: sensor, propriedade, entidade de interesse, observação (ou sensoreamento), resultado entre outros.

[3]https://www.w3.org/2003/01/geo/
[4]https://developers.google.com/maps/documentation/javascript/maptypes?hl=pt-br
[5]https://www.instagram.com/developer/endpoints/

- Representação ontológica das informações de contexto: a representação realizada por *Hermes Widget* deve representar formalmente as informações de contexto sendo, neste caso, baseada em ontologias.
- Padronização do modelo de contexto: *Hermes Widget* deve permitir a representação padronizada do contexto, independente do domínio de aplicação.
- Disseminação de informações de contexto: após representar o contexto recebido da camada de aquisição, o *Hermes Widget* deve disponibilizar essas informações para demais componentes do CMS *Hermes* e/ou aplicações interessadas.

Complementando tais requisitos, este trabalho adiciona a representação ontológica de informações de localização *outdoor* (localização em coordenadas geográficas), baseada no padrão WGS84, oferecendo expressividade e interoperabilidade para sua utilização por aplicações em ambientes de IoT.

A Figura 5 apresenta a arquitetura do componente *HW Locator*. Ligado à camada de representação, o modelo ontológico adotado utiliza as ontologias SSN, DUL e WGS84 para representar a semântica dos dados de sensores e oferecer as informações interoperáveis. Como previsto no projeto de manutenabilidade do HW, apenas o serviço de representação necessita ser alterado, sendo os demais serviços, de configuração, comunicação e notificação, herdados do projeto original.

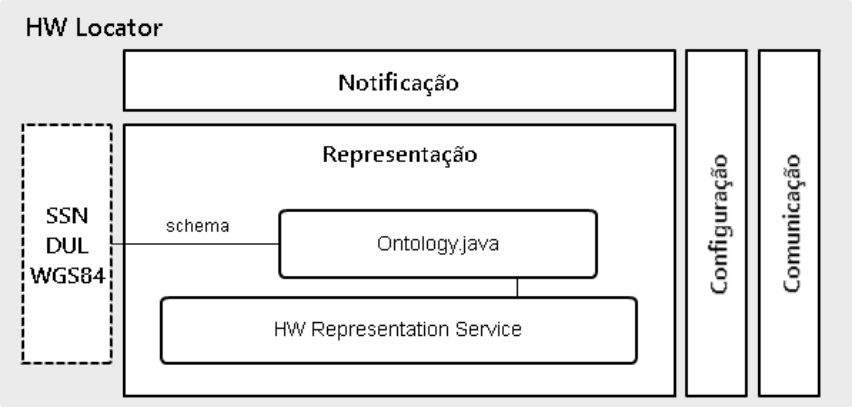


Figura 5. Arquitetura do *Hermes Widegt Locator*.

3.2. Serviço de representação de localização

O serviço de representação ontológica de localização oferecido por *HW Locator* foi desenvolvido com o objetivo de fornecer semântica e interoperabilidade para informações de localização *outdoor*. Informações de localização *outdoor* são aquelas que representam um determinado ponto no espaço, que possui latitude, longitude e altitude, referenciando a posição de uma determinada entidade.

Em ambientes IoT a informação de localização possui importância fundamental para aplicações. Como exemplo pode ser citado o cenário apresentado na introdução, que apresenta o monitoramento remoto de sinais vitais de pacientes em casa (*homecare*) e/ou em movimento. Sensores conectados à internet e monitorando pacientes enviam informações sobre seu estado para aplicações, que podem realizar serviços de acordo com as necessidades de cada paciente, tais como enviar um alarme para um profissional de saúde ou um familiar.

As informações obtidas diretamente de sensores possuem pouca expressividade e, em sua maioria, o formato em que são disponibilizadas não é interoperável. Ao serem recebidas pelo serviço de representação de *HW Locator*, para cada dado de localização é criada uma nova instância da classe *HWSensorLocator*, que recebe os parâmetros *latitude*, *longitude* e *idEntidade*, como apresentado na Figura 6. A classe *HWSensorLocator* através do seu *constructor* inicializa todas as propriedades recebidas e inicia a representação através do método herdado da classe *HermesWidgetSensorClient*, que realiza a representação das informações, utilizando como base as definições do domínio, modeladas nas ontologias SSN, DUL e WGS84.

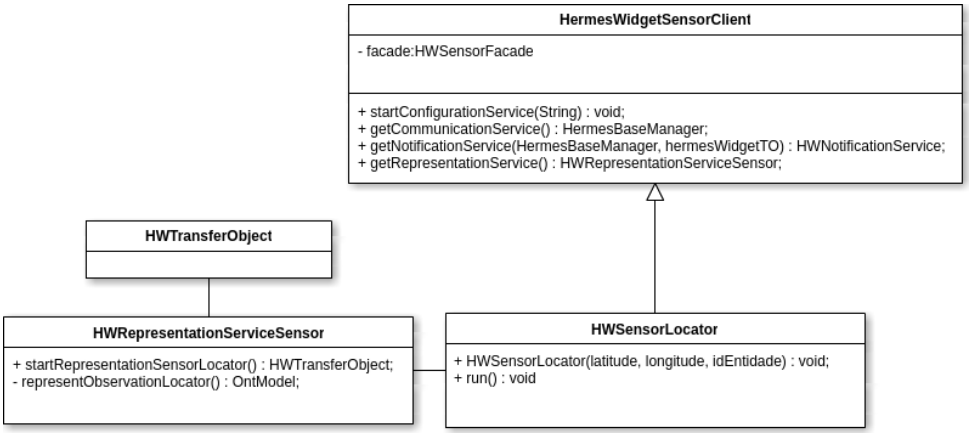


Figura 6. Diagrama de Classe do *serviço de representação de localização*

O serviço de representação gera um modelo de dados no formato *Resource Description Framework*[6] (RDF), que é a estrutura padrão para representação de informações na web semântica, sendo o seu principal requisito a interoperabilidade. O RDF utiliza URIs para identificar uma entidade e também criar uma relação entre entidades. Essas relações são denominadas triplas, onde cada tripla é composta por um sujeito, um predicado e um objeto. Um conjunto de triplas RDF constitui um grafo RDF [Klyne and Carroll 2006].

Para trabalhar com este tipo de informação semântica e interoperável, foi utilizada a API Jena [7], da Apache. Dessa forma, as informações são atribuídas a uma instância da classe *OntModel*, que gerencia as triplas do modelo RDF, para que estas sejam utilizadas em memória, enviadas para aplicações ou para que o banco de dados posso armazenar e realizar inferências na representação.

A classe *HWRepresentationServiceSensor* faz uso das ontologias *SSN*, *DUL* e *WGS84*, para que as informações de localização sejam representadas recebendo expressividade semântica suficiente para que seu entendimento seja único, independente da aplicação que a utilize. Para compreensão uniforme e interoperabilidade da informação, os dados de localização no padrão *WGS84* são representados de maneira alinhada às informações de sensoreamento no padrão SSN, através da ontologia DUL, de maneira a gerar uma informação expressiva, com sentido único e completo.

[6]https://www.w3.org/RDF/
[7]https://jena.apache.org/

A Figura 7 apresenta um exemplo básico de representação semântica interoperável de localização onde a ontologia DUL, através da propriedade *dul:hasLocaton*, interliga a informação de localização, descrita pelo indivíduo do tipo *geo:Point*, ao sensor que realiza a aferição da latitude e da longitude.

```
:sensor a ssn:Sensor ;
    dul:hasLocation [
        a geo:Point ;
        geo:latitude "31.11200108" ;
        geo:longitude "61.88699752"
] .
```

Figura 7. Estrutura de representação de localização do sensor.

Após o método *startRepresentationSensorLocator()* da classe *HWRepresentationServiceSensor* reunir todas essas informações em um RDF e retornar uma instância *OntoModel*, o componente *HW Locator* disponibiliza esse RDF para os demais componentes e aplicações que necessitam dessas informações de contexto. Essa disponibilização de informação de contexto ocorre quando o método *startRepresentationSensorLocator()* da classe *HWRepresentationServiceSensor* retorna uma instância da classe *HWTransferObject*, que encapsula os dados que serão notificados pelo serviço de representação de localização, e através do método *getNotificationService()*, herdado pela classe *HermesWidgetSensorClient*.

Voltando ao nosso exemplo do início da subseção, quando a informação de localização agregada com as informações de sinais vitais do paciente chegam às aplicações, elas analisam e decidem, que há a necessidade de uma intervenção médica. Com informações combinadas, uma equipe de salvamento em emergência pode ser acionada automaticamente para o deslocamento até a localização do paciente, ganhando em tempo e eficiência, em relação aos meios tradicionais, como uma ligação para solicitar um atendimento emergencial.

3.3. Validação do Serviço

Como dito anteriormente, os HW's não fazem nenhum tipo de aquisição de informação, dependendo de servidores externos para enviar a ele informações de sensoriamento. Como forma de validação e testes foi utilizada uma base de dados pública do *Simplemaps*[8] no formato *CSV*. Essa base de dados contendo cidades existentes no mundo, contém as seguintes informações:

- *city*, nome da cidade;
- *city_ascii*, nome da cidade sem caracteres especiais, como, por exemplo, acentos;
- *lat*, latitude global da cidade;
- *lng*, longitude global da cidade;
- *pop*, população média;
- *country*, país que a cidade está localizada;
- *iso2*, abreviação do nome do país com dois caracteres;
- *iso3*, abreviação do nome do país com três caracteres;
- *province*, província ou estado que a cidade está localizada;

[8]http://simplemaps.com/static/demos/resources/world-cities/world_cities.csv

Uma aplicação em Java, realiza a leitura das informações do arquivo *CSV* e simula um sensor, enviando para o *HW Locator* as informações de latitude, longitude como primeiro e segundo parâmetro. Ao percorrer o arquivo *CSV* a aplicação envia diferentes pontos para o serviço de representação e como forma de ligação de uma entidade a localização, é enviado de forma fixa a *String "idEntidade"* como terceiro parâmetro simulando o *id* da entidade que possui a localização representada.

A Figura 8 apresenta um exemplo de representação de localização de uma entidade, produzido pelo *HWLocator*.

```
1 @prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
2 @prefix ns0: <http://purl.oclc.org/NET/ssnx/ssn#> .
3 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
4 @prefix ns1: <http://www.loa-cnr.it/ontologies/DUL.owl#> .
5
6 geo:localizacao_entidade-b9fc62b4-fbf5-461e-8fde-d83cafcc1dd6
7   geo:long 6.209682e+1 ;
8   geo:lat 3.239173e+1 ;
9   a geo:Point .
10
11<http://purl.oclc.org/NET/ssnx/ssn#observation-lct-49d15c36-3827-4790-8cf0-ad5ddb2a31dd>
12  ns0:observationResultTime "2016-09-03T19:25:58.809-03:00"^^xsd:dateTime ;
13  ns0:featureOfInterest ns0:idEntidade ;
14  ns0:observedProperty ns0:property-lct ;
15  ns0:observedBy ns0:sensor-L_0000001H-7d99b3b3-4934-4073-af4f-a1e317c26e7a ;
16  a ns0:Observation .
17
18ns0:sensor-L_0000001H-7d99b3b3-4934-4073-af4f-a1e317c26e7a
19  ns0:observes ns0:property-lct ;
20  ns1:hasLocation geo:localizacao_entidade-b9fc62b4-fbf5-461e-8fde-d83cafcc1dd6 ;
21  a ns0:SensingDevice .
22
23ns0:idEntidade a ns0:FeatureOfInterest .
24ns0:property-lct a ns0:Property .
```

Figura 8. Representação da localização em RDF no formato TURTLE.

Nas linhas 1 a 4 são definidos os prefixos para as URIs das ontologias e do *XMLSchema* que define tipos de dados básicos, por exemplo, tipo *Date*, *int*, *double*, entre outros. Esses prefixos são utilizados para simplificar a escrita e a legibilidade da representação. Nas linhas 6 a 9 é definido um indivíduo de *geo:Point*, classe da ontologia *WGS84*. Na linha 6 é definido um nome para a instância da classe, e na linha 7, a propriedade *geo:long* referente a longitude, seguida pela propriedade *geo:lat* referente a latitude, na linha 8.

O sensor *ns0:SensingDevice* é definido nas linhas 18 a 21. O *geo:Point* é referenciado pelo nome *geo:localizacao_entidade-b9fc62b4-fbf5-461e-8fde-d83cafcc1dd6* na linha 20, e a essa localização é adicionada a propriedade *ns1:hasLocation* da classe *ns0:SensingDevice*. A propriedade sensoreada pelo sensor é adicionada na linha 24, e a propriedade *ns1:hasLocation*, definida pela ontologia *DUL*, interliga o sensor SSN e a localização WGS84.

Por último a junção de todas as propriedades definidas, são representadas nas linhas 11 a 16. Na linha 11 é definido o nome, igualmente definidas pelas outras instâncias. Para informações de momento de representação da informação na linha 12 é utilizado a definição do *XMLSchema* definindo uma data para a propriedade *ns0:observationResultTime*. Na linha 13 é informada a característica de interesse, representado pela propriedade *ns0:featureOfInterest* e seu valor é definido na linha 23,

sendo que essa propriedade é o id da entidade que está sendo observada. Na linha 14 é adicionado a propriedade *ns0:observedProperty* informando qual propriedade essa *ns0:Observation* está observando. Na linha 15 é adicionado quem observa essa propriedade através do *ns0:observedBy* e seu valor foi definido na linha 18 com o nome *ns0:sensor-L_0000001H-7d99b3b3-4934-4073-af4f-a1e317c26e7a*. E por último a linha 16 que é especificado de qual classe são as propriedades.

Após a criação da representação, o *HW Locator* envia o resultado através do serviço de comunicação para os demais componentes e aplicações, que vão utilizá-la conforme suas necessidades. A utilização da ontologia SSN, que implementa padrão Estímulo-Sensor-Observação, contribui facilitando o uso das informações compartilhadas, fazendo com que qualquer aplicação que utilize o contexto gerado pelo *Hermes Widget* trabalhe de forma análoga [Veiga and Bulcão Neto 2016].

O *HW Locator* segue uma estrutura idêntica a de qualquer outro HW, facilitando que outros componentes do CMS *Hermes* possam agregar a informação de localização a representações de outros domínios. Para que seja identificado que determinada representação é de uma entidade como, por exemplo a de paciente demonstrada por [Veiga and Bulcão Neto 2016], é utilizado o *id* do paciente, ou seja, uma entidade da base de dados pública adotada por ele. Esse *id* da entidade é utilizado para ligar um paciente aos sinais vitais, recebidos e representados por cada HW. Quando alguma aplicação necessita de informações de localização e sinais vitais o componente *Hermes Aggregator* agrega essas informações utilizando o id da entidade entregue no início da representação feita pelos HW's, promovendo a interoperabilidade entre qualquer aplicação que utilize essas informações.

4. Trabalhos Relacionados

Essa seção realiza a comparação de *HW Locator* com outros trabalhos que realizam representação de localização. Como comparativo, são verificadas quais ontologias ou técnica foi utilizada e como é realizado o serviço de representação de informações.

[Veiga and Bulcão Neto 2015] reúsam as ontologias espaço-temporal *OWL-Time* e de localização WGS84 para realizar o desenvolvimento de uma ontologia própria chamada de *OntoBus*. Essa ontologia tem como principal objetivo realizar o gerenciamento de informações de redes de transporte coletivo, onde uma linha de ônibus é construída por vários pontos espaciais que possuem latitude e longitude. Para saber quando determinado ônibus passa pelo ponto de ônibus, ele simula um sensor de localização que fica enviando sua localização e tempo para um servidor que realiza a representação das informações e armazena para consultas futuras.

[Silva Júnior et al. 2015] desenvolveram um sistema denominado de *Search Class*, a partir de um aplicativo instalado em qualquer dispositivo móvel que possua um sensor GPS, direcionar o usuário no campus da universidade, mostrando o caminho mais próximo para ele chegar em sua sala. Essa aplicação utiliza coordenadas GPS e funcionalidades da API *Google Maps V2*. [Silva Júnior et al. 2015] desenvolveram um sistema capaz de calcular e recalcular rotas rápidas, conseguindo encontrar salas dentro de construções definindo um mapeamento. Essas informações são armazenadas em um banco de dados da universidade, não foi descrito pelo autor a forma de armazenamento das informações. Não podendo concluir se as informações armazenadas possuem con-

texto suficiente, compartilhável e interoperável com outros aplicativos ou sistemas, sendo necessário ter conhecimento aprofundado da estrutura do banco de dados.

[Gularte et al. 2016] propõe dois aplicativos denominados de ônibus Online Robot e ônibus Online User, respectivamente suas funções são: *i)* monitoramento dos ônibus através de um dispositivo com suporte a plataforma android conectado a internet e GPS, enviando coordenadas geoespaciais do veículo com um intervalo de tempo de cinco segundos, para uma base de dados remota; *ii)* destinado a usuários do transporte coletivo, com o aplicativo instalado no dispositivo, os usuários recebem do ônibus monitorado, informações de distância e tempo estimado de chegada até o ponto selecionado. O armazenamento das informações coletadas pelos sensores são armazenadas remotamente no banco de dados relacional MySQL. O diagrama de classes afirma que as informações armazenadas possuem pouca expressividade dificultado o compartilhamento e interoperabilidade.

A proposta de [Franco and Perozzo 2015] visa a criação de um sistema para gerenciamento de táxi, sub-dividido em três partes: *i)* gerenciador do taxista; *ii)* gerenciador do passageiro; e *iii)* gerenciador web. Sua proposta é de usuários solicitarem táxi de forma personalizada, contendo ar-condicionado, cadeira de bebe onde é compartilhada a geolocalização de veículos e passageiros. Através do gerenciador do passageiro ele pode solicitar um táxi, o taxista a partir do gerenciador do taxista pode visualizar a localização do passageiro e dizer se quer ou não passageiro e do gerenciador web é possível saber a localização de todos os táxis da frota. Sua proposta é parecida com outros aplicativos de gerenciador de táxi existentes. De acordo com o modelo entidade relacionamento do banco de dados, o sistema não faz uso de ontologias para representação, armazenando a informação com pouca expressividade.

Em relação ao trabalho de [Veiga and Bulcão Neto 2015], o serviço de representação de localização *HW Locator* utiliza a ontologia SSN para representação e estruturação semântica de informações de sensores. E em relação aos outros três trabalhos apresentados nenhum utiliza ontologia para representação de informação de localização armazenando informação com um contexto baixo, sem expressividade forçando que os sistemas, que necessitem da informações de localização, conheçam bem a estrutura armazenada. Por realizar representações expressivas os sistemas e aplicações não precisam conhecer profundamente a estrutura do banco de dados, necessitando apenas conhecer as ontologias utilizadas para a representação [Klyne and Carroll 2006].

O serviço de representação de localização descrito também foi criado de forma desacoplada, podendo ser utilizado em outra infraestrutura que tenha a mesma base de comunicação que o *Hermes*. Por conta desse desacoplamento aplicações não precisam se preocupar em processar informações, somente recebendo e utilizando essas informações para sua necessidade. Para que o serviço de representação de localização fosse criado se teve um esforço menor em relação ao desenvolvimento completo, pois grande parte foi reusado.

5. Conclusão

Este artigo apresentou *HW Locator*, um serviço de representação de informações de localização, um dos mais importantes tipos de informações em ambientes IoT. Inicialmente foi apresentado um dos cenários de aplicação para este serviço, onde equipes

médicas utilizando diversos aplicativos diferentes acessam as informações de localização de um paciente, que são interpretadas de maneira uniforme. O serviço de representação de localização proposto produz um modelo de contexto expressivo e em formato que permite a interoperabilidade entre os interessados na informação, atendendo a este e a outros cenários de IoT.

O trabalho se limitou a realizar a representação de informações *outdoor*, onde podem ser representados apenas pontos ou coordenadas em um mapa, definidos por *latitude*, *longitude* e *altitude*. Porém, este trabalho pode ser estendido para realizar também representações *indoor*, utilizando os padrões e ontologias existentes para este tipo de informação de localização.

Como trabalho futuro, pretende-se realizar a representação das informações de localização *indoor*, onde há a necessidade de identificar, por exemplo, entidades situadas dentro de edificações, as quais não podem ser localizadas com precisão pelo sistema de navegação por satélite (GPS).

Referências

Bastos, A. B. et al. (2013). Uma abordagem ontológica baseada em informações de contexto para representação de conhecimento de monitoramento de sinais vitais humanos.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Phuoc, D. L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., and Taylor, K. (2012). The {SSN} ontology of the {W3C} semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17:25 – 32.

Franco, J. R. and Perozzo, R. F. (2015). Sistema para localização e gerenciamento de táxis. *Disciplinarum Scientia— Naturais e Tecnológicas*, 16(1):95–107.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). *Sweetening Ontologies with DOLCE*, pages 166–181. Springer Berlin Heidelberg, Berlin, Heidelberg.

Gularte, Á., Ribeiro, V., and Silveira, S. (2016). Um protótipo para monitoramento em tempo real do transporte público de porto alegre - rs por gps (global positioning system). *RCT - Revista de Ciência e Tecnologia*, 2(2).

Klyne, G. and Carroll, J. J. (2006). Resource Description Framework (RDF): Concepts and Abstract Syntax.

Kumar, M. (1988). World geodetic system 1984: A modern and accurate global reference frame. *Marine Geodesy*, 12(2):117–126.

Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *IEEE Communications Surveys Tutorials*, 16(1):414–454.

Silva Júnior, G. A., Soares, A. C. P., Pereira, M. H. R., and Maia, E. H. B. (2015). Search class: Sistema para gerenciamento de rotas em universidades por meio de módulos gps e pontos de localização não mapeados pelo google. *e-xacta*, 8(1):117–129.

Veiga, E. F. and Bulcão Neto, R. F. (2015). An ontological approach for information management of public transport networks. In *Proceedings of the Annual Conference on Brazilian Symposium on Information Systems: Information Systems: A Computer Socio-Technical Perspective - Volume 1*, SBSI 2015, pages 66:493–66:500, Porto Alegre, Brazil, Brazil. Brazilian Computer Society.

Veiga, E. F. and Bulcão Neto, R. F. (2016). Um Serviço de Representação Ontológica de Contexto Baseada no Padrão de Projeto Estímulo-Sensor-Observação. *VIII Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP)*, pages 1–10.

Veiga, E. F., Maranhão, G. M., and Bulcão Neto, R. F. (2016). Development and scalability evaluation of an ontology-based context representation service. *IEEE Latin America Transactions*, 14(3):1372–1379.

Veiga, E. F., Maranhão, G. M., and Bulcão Neto, R. F. (2014). Apoio ao desenvolvimento de aplicações de tempo real sensíveis a contexto semântico. *IX WTD do XX Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 1–4.

Defendendo Aplicações Web: Mapeando as Principais Vulnerabilidades e seus Controles de Segurança

Murilo Carvalho Libaneo[1], Sérgio Teixeira Carvalho[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.001-970 – Goiânia – GO – Brasil

mclibaneo@gmail.com, sergio@inf.ufg.br

Abstract. *The security systems aims to protect the information as to its confidentiality, integrity and availability, so defend systems of threats has become a key requirement. The objective of this paper is to assist in the development of safety-oriented software, through a mapping controls for defense and prevention against the top ten vulnerabilities listed in the report of the Open Web Application Security Project (OWAP). The OWASP initiative was used as reference due to the objectivity of their materials and their philosophy according to the free software movement.*

Resumo. *A segurança de sistemas tem como objetivo proteger as informações quanto à sua confidencialidade, disponibilidade e integridade, logo, defender os sistemas das ameaças tornou-se um requisito fundamental. O objetivo deste trabalho é auxiliar no desenvolvimento de software orientado a segurança, por meio de um mapeamento de controles, para a defesa e prevenção contra as dez maiores vulnerabilidades listadas no relatório da Open Web Application Security Project (OWAP). A iniciativa da OWASP foi utilizada como referência, devido à objetividade de seus materiais e por sua filosofia em consonância com o movimento de software livre.*

1. Introdução

A segurança da informação é uma área que vem ganhando destaque a cada ano, principalmente por conta das mudanças tecnológicas. Dentre estas mudanças nota-se o surgimento de uma internet aprimorada para fornecer sistemas complexos e distribuídos, e a ampla adoção de aplicações móveis – presentes em computadores, *smartphones*, dispositivos vestíveis, entre outros.

Em consequência das mudanças tecnológicas surgiram novas ameaças, como é o caso do *Cross Site Scripting* (XSS) para as aplicações *web* [Barbosa *et al.* 2015]. A situação agrava-se, pois os antigos métodos de ataque e invasão passaram por um processo de sofisticação, enquanto que a engenharia de *software*, apesar dos avanços de padrões de desenvolvimento e de qualidade de *software*, ainda não atingiu um nível de maturidade com a segurança e confiabilidade como metas de projeto [Stalling *et al.* 2014].

Segundo dados estatísticos do CERT.br[1], no ano de 2015 foram registradas 722.205 notificações de incidentes em segurança. Apesar de ser um número menor em relação à quantidade de notificações de 2014 (1.047.031), nota-se, especificamente, um aumento de 128% nos ataques aos servidores *web*. De acordo com o CERT.br, os atacantes exploram vulnerabilidades em aplicações *web*, e, a partir de então, comprometem a segurança do sistema.

Neste cenário, é possível perceber a crescente dificuldade em garantir os objetivos fundamentais da segurança da informação: confiabilidade, integridade e disponibilidade [Stallings *et al.* 2014]. Felizmente, foram desenvolvidas abordagens e técnicas que permitissem mitigar os problemas em segurança. Este trabalho apresenta uma abordagem para a defesa proativa, contra as principais vulnerabilidades de aplicações *web*, por meio de controles de segurança.

São vários os tipos de vulnerabilidades que uma aplicação *web* possui. O relatório da Open Web Application Security Project (OWASP) [Machry 2013], traz os dez maiores riscos de segurança em aplicações *web*. Este relatório, além de classificar os principais riscos de segurança, possui também técnicas de proteção contra as ameaças e orientações para as organizações vítimas de ataques.

Com o propósito de simplificar o emprego da segurança em *software*, este estudo traz um mapeamento das formas de defesa contra as dez principais vulnerabilidades, utilizando como material norteador o Guia de Desenvolvimento da OWASP [Stock 2005] e sua versão atual [Baan *et al.* 2015], o *Proactive Controls Project* [Anton *et al* 2016], o *Application Security Verification Standard* [Manico *et al.* 2016], além de conhecimentos adquiridos no portal da iniciativa OWASP[2].

O artigo está organizado da seguinte forma: a Seção 2 aborda os conceitos de desenvolvimento de *software* orientado a segurança; a Seção 3 justifica o uso do guia de desenvolvimento da OWASP, e seus anexos, além de explicá-lo; na Seção 4 são apresentadas dez vulnerabilidades e seus respectivos controles de segurança; e por fim, a Seção 5 concentra as considerações finais do trabalho.

2. Desenvolvimento de Software Orientado a Segurança

Embora segurança da informação e segurança de sistemas sejam conceitos diferentes, eles compartilham a mesma base teórica. A segurança de sistemas é uma área que auxilia a segurança da informação a garantir os princípios da confiabilidade, integridade e disponibilidade.

O desenvolvimento de *software* orientado a segurança é uma das vertentes, dentro da segurança de sistemas, responsável pela proteção contra vulnerabilidades internas e externas aos sistemas e promove o uso de diretrizes de segurança em todo o seu ciclo de vida. O termo sistemas foi empregado como uma forma abrangente para exemplificar o *software*, as aplicações *web* e outros. Porém, em Pressman (2011) é possível compreender a especificidade destes termos. Neste artigo evitou-se usar o termo *software* seguro, pois remete a ideia de um produto protegido cem por cento contra as ameaças, e segundo Sommerville (2011) esta ideia é errônea.

[1] As estatísticas completas dos incidentes reportados podem ser encontradas no portal do CERT.br: http://www.cert.br/stats/incidentes/

[2] *Site* do Projeto OWASP: https://www.owasp.org/

Para Mcgraw (2006) a segurança de *software* é a prática de construir *softwares* que resistam a ataques de forma proativa. Define, ainda, a necessidade de uma abordagem holística, combinada com a filosofia de polaridades, no qual a realidade se constrói a partir da ação (ataque) e reação (defesa). Como exemplo, o autor cita a prática de realização de testes contra a segurança de um *software*. Estes testes, muitas vezes invasivos, são considerados atividades de chapéu preto (*black hat*), pois são destrutíveis e comprometem o funcionamento padrão do *software*.

Por outro lado, as atividades de chapéu branco (*white* hat) são aquelas que auxiliam na segurança e no funcionamento adequado de um *software*. Um exemplo são as boas práticas de programação, os controles de segurança e afins. Estas duas práticas, chapéu preto e chapéu branco, contribuem, em conjunto, para o desenvolvimento mais seguro de um *software*.

Neste sentido, este estudo empregou os conceitos das práticas de chapéu branco na defesa das aplicações *web*, uma vez que não foi necessária a realização de técnicas de invasão. Utilizou-se os controles de segurança em vulnerabilidades já conhecidas, porém a prática de chapéu preto é eficaz na descoberta de possíveis falhas de segurança remanescentes.

Promover a segurança de sistemas é um processo complexo e perpassa por uma série de desafios. Entre estes problemas está o fato do tempo extra gasto no desenvolvimento de uma solução de segurança, e a falta de definições claras dos requisitos de segurança [Stallings *et al.* 2014].

O uso de diretrizes de segurança, como, por exemplo, o Guia de Desenvolvimento da OWASP, descomplica alguns destes problemas, pois contém estruturas já definidas que agilizam o desenvolvimento. É o caso dos controles de segurança, descritos neste trabalho. Outra característica destas diretrizes, é o fato de empregarem a segurança em todo o ciclo de vida do *software*, como no caso do Microsoft Security Development Lifecycle (MSDL). O MSDL é uma diretriz que facilita o trabalho de gerentes de projetos, com ênfase em minimizar as vulnerabilidades em *software* e também introduzir a "privacidade em todas as fases do processo de desenvolvimento" [Microsoft 2010].

3. O Guia de Desenvolvimento da OWASP

O Open Web Application Security Project (OWASP) é uma comunidade aberta, sem fins lucrativos, com o foco em descobrir e reduzir os problemas de segurança em *software*. Ela reúne vários projetos com o mesmo objetivo e é mantida por profissionais e entidades de todo o mundo. Um destes projetos é o OWASP *Development Guide*, ou Guia de Desenvolvimento OWASP (GDO), que se concentra na parte de escrita de códigos legíveis e seguros, e alinha-se com o *Application Security Verification Standard* e o OWASP *Testing Guide*[3].

O GDO não é uma bala de prata contra todos os riscos que ameaçam a segurança de um *software* [Stock 2005]. Ele foca nas boas práticas de codificação e nos princípios de segurança, mas aconselha fortemente a utilização do gerenciamento de riscos, a fim de reduzir os custos e o tempo de desenvolvimento. Quanto mais cedo descobrir e identificar as principais ameaças de um projeto de *software* (análise de risco), menor é a

[3] É um projeto que inclui um *framework* adaptável para realização de testes em aplicações *web* e serviços *web:* https://www.owasp.org/index.php/OWASP_Testing_Project

probabilidade de introduzir vulnerabilidades prejudiciais à segurança do produto [Baan *et al.* 2016], [Mcgraw 2006] e [Howard *et al.* 2006].

Na Tabela 1 são apresentados os princípios da arquitetura de segurança de sistema. Os controles de segurança englobam na prática muito destes princípios e serão conferidos na Seção 4.

Tabela 1. Princípios da Arquitetura de Segurança. Adaptado de [Baan *et al.* 2016]

Princípios da Arquitetura de Segurança	
Defesa em profundidade	Significa ter diferentes camadas de proteção ao longo do *software*. Na prática, é manter diferentes tipos de mecanismos de defesa. Como exemplo, as validações em *Javascript* no lado do cliente e as mesmas validações no lado do servidor.
Falha Segura	Relacionado à resiliência de *software*. Quando ocorrer uma falha, o *software* deve se recuperar de forma a voltar a um estado seguro antes da ocorrência da falha.
Privilégios Mínimos	Dentro da aplicação, um indivíduo ou processo, deve ter o mínimo de direito de acesso (privilégio), necessário para realizar um determinado procedimento.
Separação de Deveres	Este princípio propõe que um procedimento ou ação dentro do *software*, depende de duas ou mais condições para ser concluído. Um cadastro, por exemplo, necessita de o usuário realizar *login* (com privilégios de cadastro), e, posteriormente ,confirmar as informações.
Economia de Mecanismos	Alinha-se com a filosofia *KISS (Keep It Simple, Stupid)* e defende a simplicidade da arquitetura e da implementação de *software*. Quanto menor a complexidade de um *software*, menor é a possibilidade da existência de vulnerabilidades internas
Mediação Completa	Qualquer ação que envolva dados sensíveis deve ser verificada e validada quanto à autenticação e ao acesso. Na prática, significa verificar, sempre que necessário, o *login* e o privilégio de um usuário.
Arquitetura Aberta	Este princípio defende que os detalhes de implementação de uma arquitetura sejam independentes da própria arquitetura, e que, salvos os casos em que a segurança de um *software* necessita de uma arquitetura fechada, a segurança da aplicação não deve depender da arquitetura em si.
Mecanismos Comuns Mínimos	Propõe o não compartilhamento de mecanismos (funcionalidades de *software*) entre usuários ou processos, se estes usuários ou processos tiverem privilégios distintos.
Aceitabilidade Psicológica	Usar um *software* deve ser o mais fácil e simples possível. Se os mecanismos de segurança de um *software* dificultarem o trabalho de um usuário, é grande a possibilidade deste usuário contornar indevidamente estes mecanismos e abrir brechas dentro do *software*. O princípio propõe que a usabilidade e transparência dos mecanismos de segurança, são essenciais para a efetividade da segurança da aplicação.
O Elo Mais Frágil	Este princípio propõe identificar e reparar o elo mais frágil, aquele que mais facilmente comprometerá a segurança do *software*, seja ele um código, um serviço ou o usuário.
Reuso de Componentes Existentes	O princípio defende que ao utilizar componentes seguros, já existentes em outras partes do *software*, evita a introdução de novas vulnerabilidades e evita o aumento dos vetores de ataques.

O Guia de Desenvolvimento da OWASP é uma diretriz específica para a codificação segura das aplicações. Apesar de existirem outros processos e guias, como o MSDL e o *Framework* da *Secure Software Foundation*[4], o GDO foi escolhido principalmente pela sua filosofia inteiramente aberta e por ser referência em diversos materiais de relevância, como livros, artigos, legislações e normas[5]. Outro importante fator é o seu formato simplificado. Ao separar em anexos os assuntos distintos, torna

[4] É uma iniciativa recente de empresas de segurança de software Holandesas, *site* do projeto: https://www.securesoftwarefoundation.org/index.php/about-us/

[5] Referências que utilizam o OWASP: https://www.owasp.org/index.php/Industry:Citations#National_.26_International_Legislation.2C_Standards.2C_Guidelines.2C_Committees_and_Industry_Codes_of_Practice

mais fácil tanto a leitura quanto a utilização de forma prática dos controles de segurança.

4. As Vulnerabilidades das Aplicações Web e os Controles de Segurança

Como muitas das vulnerabilidades conhecidas são recorrentes em vários projetos de aplicações *web*, a OWASP criou um documento contemplando oito conjuntos de dados baseados em sete empresas especializadas em segurança [Machry 2013]. Estes dados resultaram em uma lista com as dez principais vulnerabilidades encontradas em aplicações *web*. Cada vulnerabilidade possui um risco associado e atributos específicos para cada empresa, como, por exemplo, os agentes de ameaça e os impactos no negócio. Este artigo preocupou-se especificamente no tratamento destas vulnerabilidades, mapeando os controles essenciais para proteção e prevenção das aplicações *web*.

4.1. A1 – Injeção

Ataques de injeção ocorrem quando uma entrada de dados influencia acidentalmente ou deliberadamente o fluxo de execução do programa [Stallings *et al.* 2014]. O principal controle para este tipo de vulnerabilidade é a **Validação de Entradas**. A validação de entradas garante proteção contra vários tipos de ataques de injeção, entre eles o ataque de injeção SQL.

Para [Anton *et al*. 2016] a melhor alternativa para sanar os problemas de injeção SQL é o método conhecido como '*Query Parameterization'*, ou Consultas Parametrizadas. Neste método, cancela-se os efeitos dos caracteres de escape, filtrando as consultas e enviando-as separadamente dos parâmetros ao servidor do banco de dados. O uso de *frameworks* apropriados para cada linguagem de programação, como o *Hibernate*[6] para o JAVA e o *Doctrine*[7] para PHP, facilitam o trabalho de mapeamento das entidades para uma base de dados.

Em paralelo ao uso das consultas parametrizadas, Sadeghian (2013) citado por Carvalho e Júnior (2014) sugere o uso da Configuração Inteligente do Esquema de Privilégios do Banco de Dados, que consiste na criação de usuários com permissões distintas de interação com o banco, sendo permitida apenas consultas para determinadas páginas. Porém, esta técnica não impossibilita a visualização das informações do banco caso um ataque de injeção seja bem-sucedido.

Na Figura 1 é demonstrado um exemplo do uso de consultas parametrizadas utilizando a linguagem Java, as informações de 'nomeCompleto' e 'codigo' são enviadas separadamente e filtradas por meio da classe *PreparedStatement.*

```
1  //Controles de Segurança - Validação de Entradas
2  String nomeCompleto = request.getParameter("nomeCompleto");
3  int codigo = Integer.parseInt(request.getParameter("codigo"));
4  PreparedStatement pstm = conexao.PreparedStatement("UPDATE funcinarios SET nomeCompleto = ? WHERE codigo = ?");
5  pstm.setString(1, nomeCompleto);
6  pstm.setInt(2, codigo);
```

Figura 1. Query Parameterization em Java. Adaptado de [Anton *et al.* 2016]

Toda e qualquer entrada deve ser verificada e validada. Este processo pode ocorrer tanto no navegador (*front end)* da aplicação quanto no servidor (*back end)*, sendo mais aconselhável que as validações aconteçam no servidor, pois o atacante pode burlar os mecanismos de validação no navegador.

6 *Site* do Projeto *Hibernate*: http://hibernate.org/

7 *Site* do Projeto *Doctrine*: http://www.doctrine-project.org/

As validações podem ser sintáticas, quando se valida a forma do dado, ou semânticas, quando se valida o conteúdo do dado dentro do contexto da aplicação. As validações sintáticas podem empregar o método de lista branca ou lista preta, sendo este menos recomendável, pois é mais permissivo. O método de lista preta assume que todas as entradas são permitidas, menos aquelas já conhecidas como sendo maliciosas. O outro método é mais rigoroso, bloqueando todas as entradas e permitindo somente aquelas que se encontram em um dicionário, ou que atendam a um padrão de uma expressão regular, ou seja, "um modelo composto por uma sequência de caracteres que descreve variantes de entradas permitidas." [Stallings *et al.* 2014].

4.2. A2 – Quebra de Autenticação e Gerenciamento de Sessão

Autenticação é "o processo de verificação de uma entidade alegada por ou para uma entidade de sistema" [Stallings *et al.* 2014] e consiste em duas etapas: a etapa de criação de uma identificação e a etapa de verificação da identidade no sistema. Esta vulnerabilidade está relacionada ao corrompimento da identificação de um usuário, geralmente quando há extravio de senhas, *tokens* e identificadores por parte do atacante, e este se passa por um usuário legítimo. A principal defesa contra esta vulnerabilidade é o **Gerenciamento de Sessão.**

Um gerenciamento de sessão adequado, conforme explicado por [Anton *et al.* 2016], possui as seguintes características: a) Utiliza autenticação de múltiplos fatores, conforme os meios de autenticação conhecidos na literatura: algo que o indivíduo conhece ou sabe (senha); algo que o indivíduo possui (*token*); algo que o indivíduo é (digital) e algo que o indivíduo faz (assinatura); b) Implementa segurança no armazenamento de senhas, através do uso de criptografia e banco de dados confiáveis; c) Fornece segurança no mecanismo de recuperação de senha, como, por exemplo, uma pergunta secreta cuja resposta somente o usuário conhece; d) Gerência de sessão conforme o uso do usuário, com função de encerrar através de um tempo de expiração; e) Possui domínio de toda transação do usuário com o sistema, com controles adequados de reconexão, e em casos sensíveis, por exemplo, mudança de senha, exigir autenticação novamente.

O *Spring Security*[8] é um *framework* para a plataforma JAVA que encapsula várias destas características. Conforme a Figura 2 é possível verificar a simplicidade de codificar uma classe de configuração para o uso do *framework.*

```
1  //Controles de Segurança - Gerenciamento de Sessão com Spring Security
2  package org.springframework.security.samples.config;
3
4  import org.springframework.beans.factory.annotation.Autowired;
5  import org.springframework.security.config.annotation.authentication.builders.AuthenticationManagerBuilder;
6  import org.springframework.security.config.annotation.web.configuration.*;
7
8  @EnableWebSecurity
9  public class SecurityConfig{
10
11     @Autowired
12     public void configureGlobal(AuthenticationManagerBuilder auth) throws Exception{
13         auth.inMemoryAuthentication().withUser("user").password("password").roles("USER");
14     }
15 }
16
```

Figura 2. Classe de Configuração Spring Security. Adaptado de [Winch 2016]

[8] *Site* do Projeto Spring Security: http://projects.spring.io/spring-security/

4.3. A3 – Cross-Site Scripting (XSS)

O ataque XSS ocorre quando uma entrada contendo um *script* malicioso é enviada ao navegador sem as devidas validações. É o caso de fóruns ou sites, com conteúdos dinâmicos, que permitem receber uma entrada de usuário e inseri-las no conteúdo do próprio site. Os resultados deste ataque incluem: desfiguração do site; redirecionamentos mal intencionados; roubo de sessões do usuário, entre outros impactos negativos.

O principal controle para esta vulnerabilidade é a **Validação de Entradas,** mencionado no item 4.1. Porém, para obter uma camada a mais de proteção pode-se aplicar a **Codificação de Saída,** que consiste em traduzir os caracteres especiais (“, ‘,\,/,|,$,<,>), em uma representação que não ofereça riscos ao interpretador da aplicação. Desta forma, quando um usuário inserir um dado contendo algum caractere especial, ele conseguirá visualizar o conteúdo inserido, sem que este ofereça risco à aplicação. Os codificadores de saída mais comuns são: *AntiXSS Library*[9], *HTML Entity Encoding*[10]**,** *JavaScript Encoding*[11].

4.4. A4 – Referência insegura e direta a objetos

Esta vulnerabilidade ocorre quando um objeto, seja um arquivo, diretório ou registro de banco de dados, não possui a devida proteção contra acessos desautorizados. Por exemplo, um arquivo passado como parâmetro de forma direta na URL

```
http://www.exemplo.com.br/pagamentos.php?arq=funcionarioA.csv
```

pode ser facilmente alterada para

```
http://www.exemplo.com.br/pagamentos.php?arq=funcionarioB.csv
```

Uma das formas de resolver esta vulnerabilidade é utilizando referências indiretas a objetos por usuário ou sessão [Machry 2016]. Por exemplo, utilizando uma tabela de mapeamentos para cada objeto permitido para o usuário. Outros mecanismos de segurança recomendáveis para evitar este problema são a **Proteção de Dados** e o **Controle de Acesso**. Estes mecanismos são detalhados nas Seções 4.6 e 4.7 respectivamente.

4.5. A5 – Configuração Incorreta de Segurança

Ao iniciar um projeto de *software* é comum aos desenvolvedores, até mesmo por economia de recursos e tempo, utilizar configurações padrões de suas ferramentas. Senhas padrões, configurações mais permissivas, logs excessivamente informativos, são algumas das características que, se não forem tratadas, podem acarretar em sérios riscos de segurança para a aplicação.

O principal controle para evitar esta vulnerabilidade é a **Configuração de Segurança.** Um processo de *hardening*[12] deve ser definido a fim de impedir que

[9] *Site* da biblioteca: https://www.microsoft.com/en-us/download/details.aspx?id=28589

[10] Uma lista com codificadores de saída de caracteres especiais utilizando *HTML Entity Encoding*, pode ser conferida em: http://www.w3schools.com/html/html_entities.asp

[11] Funções de codificação de saída em *JavaScript*: http://www.w3schools.com/jsref/jsref_encodeuri.asp

[12] Processo de reduzir vulnerabilidades de superfície: https://en.wikipedia.org/wiki/Hardening_(computing)

exposições desnecessárias e falhas de segurança se propaguem no sistema e subjacentes [Stock 2015]. O controle define ainda que as bibliotecas e componentes devam estar atualizados e seguramente configurados, e que as comunicações sejam criptografadas e autenticadas. Define ainda que o ambiente de desenvolvimento deva estar isolado do ambiente de produção, a fim de evitar possíveis brechas de segurança. Tanto o ambiente de produção quanto o de desenvolvimento devem ser verificados e validados constantemente por um administrador de segurança autorizado.

4.6. A6 – Exposição de Dados Sensíveis

O objetivo da segurança de sistema é a proteção das informações quanto à disponibilidade, integridade e confidencialidade [NIST 1995 *apud* Stallings 2014]. Porém, ficaria inviável garantir a segurança de todos os dados. É necessário criar uma política que determine quais dados são mais relevantes (sensíveis) para a organização, como, por exemplo, número de cartão de crédito para as aplicações financeiras, registros médicos para as aplicações hospitalares, entre outras modalidades de negócios.

Definidos quais são os dados sensíveis da organização, o principal controle para evitar esta vulnerabilidade é a **Proteção de Dados.** Este controle encapsula outros controles de segurança e possui os seguintes princípios:

i. Criptografar todos os dados em trânsito: este princípio faz parte do controle de **Comunicação Segura** e define o uso de criptografia na comunicação entre sistemas, sempre que se fizer necessária a proteção da comunicação entre o navegador do cliente e o servidor da aplicação, e deste com os servidores de banco de dados, de e-mail, *web services*, entre outros. Embora existam muitos métodos de criptografia de comunicação, o mais conhecido é o *Transport Layer Security* (TSL).
ii. Criptografar dados sensíveis em repouso: a **Criptografia** é um dos principais controles de segurança na proteção dos dados. Este princípio estabelece o uso de criptografia em todos os dados sensíveis que não estão sendo utilizados. É importante criptografar tanto os dados quanto a chave privada da criptografia e manter estes elementos em diferentes locais. Os módulos de criptografia devem obedecer ao princípio da falha segura e os erros devem ser totalmente gerenciáveis. Outro fator relevante para o sucesso da criptografia é a utilização de algoritmos fortes e reconhecidos pelo padrão FIPS 140-2[13] ou similares.
iii. Proteger dados sensíveis em processamento: proteger os dados em uso, bem como, resolver a vulnerabilidade descrita na Seção 4.4, envolve a utilização deste princípio. Este princípio defende a não utilização das funcionalidades de *cache* e autocompletar, no lado do navegador. Inclui ainda verificar se os dados são enviados através do corpo ou cabeçalho do protocolo HTTP e não via parâmetros visíveis, e evitar uso de referências inseguras a objetos.
iv. Limpar dados sensíveis da memória: o último princípio defende o não armazenamento de dados sensíveis, desnecessários, e o descarte de forma segura, por meio da sobrescrita da memória com zeros.

[13] É um padrão usado para validar módulos de criptografia. Foi publicado em 2001 pela *Federal Information Processing Standard*. Mas detalhes em: https://en.wikipedia.org/wiki/FIPS_140-2

4.7. A7 – Falta de Função para Controle do Nível de Acesso

Nível de acesso é a hierarquia de papéis dentro de um sistema e determina quem pode interagir com o quê. Esta vulnerabilidade está relacionada ao fato de uma função de controle de nível de acesso não ter sido configurada adequadamente ou possuir falhas em sua arquitetura, permitindo aos agentes do ataque o acesso a uma informação sem a devida autorização.

O uso adequado de um **Controle de Acesso** é primordial em aplicações com múltiplos usuários. Anton (2016) define que controle de acesso é o processo de autorização, no qual uma requisição para uma determinada função ou recurso deve ser permitida ou negada.

Este controle estabelece uma completa mediação para as funções e recursos do sistema, isto é, cada requisição a uma função ou recurso deve possuir um método autenticador. Uma boa prática consiste em negar acesso a toda função que ainda não foi configurada pelo controlador de acesso. Seguindo esta linha de pensamento, o controle utiliza o princípio de privilégios mínimos, estabelecendo que usuários novos sejam criados com acesso restrito para a maioria das funcionalidades da aplicação. Importante lembrar que validações na camada do cliente, como, por exemplo, esconder um botão de uma funcionalidade que deveria ser acessada somente por um grupo de usuários, não é aconselhável, apesar de acrescentar uma camada a mais de segurança. O correto, e mais seguro, é sempre validar a autenticação no servidor da aplicação.

Mecanismos de controle de acesso são muitos complexos, sendo que criá-los demanda recursos. Além disso, podem ser falhos e não seguros. Novamente, o uso de um *framework* confiável é uma boa alternativa. O *Apache Shiro*[14] é um *framework* de segurança para linguagem JAVA e contempla às áreas de autenticação, criptografia, autorização, gerenciamento de sessões, entre outras. O *Spring Security* é outra ferramenta que vem com estruturas pré-definidas que auxiliam no desenvolvimento.

4.8. A8 – Cross-Site Request Forgery (CSRF)

Este ataque ocorre quando um usuário de uma aplicação realiza uma requisição não intencionada, e, portanto, não legítima à aplicação vulnerável. Por exemplo, quando o usuário, vítima de um ataque de engenharia social, clica em um *link* falso ou abre uma imagem contendo um *script* malicioso. Este *link* ou *script* contém uma requisição forjada, porém com todos os parâmetros adequados, como *cookies* e informações de autenticação, a uma aplicação *web* que irá processá-la como uma requisição legítima. Se o usuário estiver autenticado no serviço da aplicação e caso este serviço não possua controles de segurança adequados, as chances de sucesso são grandes. O objetivo deste ataque é mudar o estado do usuário, geralmente a senha ou e-mail, pois o atacante ao enviar o link, não possui condições de saber se o ataque foi bem-sucedido.

Os mecanismos de proteção incluem **Comunicação Segura** e **Controle de Acesso**, ambos já mencionados neste artigo. A comunicação entre o navegador do cliente e o servidor da aplicação deve possuir um método para validar e consistir a origem da requisição. Como, por exemplo, o uso do Padrão Sincronizador de *Token*, que gera um *token* único e aleatório, sendo inserido como campo oculto no formulário da aplicação ou via URL, e sincronizado com a sessão do usuário. No momento da requisição é comparado e validado o *token* [Wichers *et al.* 2016]. Outra forma de

[14] *Site* do Projeto Apache Shiro: http://shiro.apache.org/

prevenir esta vulnerabilidade é contar com autenticação em duas etapas e/ou o uso de *captchas*, porém estas abordagens necessitam de uma maior interação do usuário com a aplicação.

4.9. A9 – Utilização de Componentes Vulneráveis Conhecidos

Componentes são os recursos utilizados em um projeto de *software*, que auxiliam no trabalho do desenvolvedor. Podem ser considerados como componentes, as bibliotecas, *plug-ins*, os serviços *web* e afins. Muitas vezes estes componentes não possuem os controles adequados de segurança e acabam comprometendo a integridade de uma aplicação *web*.

Idealmente, não deveriam ser utilizados componentes de terceiros. O mais aconselhável é utilizar soluções próprias para o projeto e enfatizar nos recursos de *frameworks* já consagrados, como o *Spring Security*, o *Apache Shiro* e *Django Security*[15] [Anton *et al.* 2016].

Caso sejam utilizados componentes, algumas práticas do controle **Configuração de Segurança** devem ser realizadas. Entre elas, estabelecer políticas de segurança que compreenda o uso adequado destes componentes, gerenciar as versões dos componentes e suas bibliotecas, e definir frequências de atualizações.

4.10. A10 – Redirecionamentos e Encaminhamentos Inválidos

As aplicações *web* às vezes possuem em sua estrutura redirecionamentos a outros *sites* ou encaminhamentos internos. Infelizmente a construção incorreta destas funções pode resultar em brechas de segurança. Esta vulnerabilidade é utilizada para redirecionar as vítimas para *sites* de *phishing* ou *malware* [Machry 2013].

Considere, por exemplo, que a aplicação *web* utiliza a seguinte estrutura de redirecionamento na linguagem JAVA:

```
response.sendRedirect(request.getParameter(“url”);
```

Sem um controle de **Validação de Entrada**, um atacante poderia adicionar sua própria URL maliciosa:

```
http://exemplo.com.br/redireciona?url=http://sitecommalware.com
```

Geralmente o atacante utiliza um *link* extenso e camuflado, confundindo a vítima que acaba não percebendo a tentativa de *phishing*. As formas seguras de encaminhamentos foram mencionadas em [Bezold 2016], e na Figura 3 observa-se exemplos de uso de redirecionamentos diretos, em diferentes linguagens de programação:

[15] Página do projeto Django Security: https://docs.djangoproject.com/en/1.8/topics/security/

```
1  //Exemplos de Redirecionamentos mais Seguros
2
3  //JAVA
4  response.sendRedirect("http://www.sitedestino.com.br");
5
6  //PHP
7  <? php header("Location: http://www.sitedestino.com.br"); ?>
8
9  //ASP.NET
10 Response.Redirect("~/diretorio/Login.aspx");
11
12 //RAILS
13 redirect_to login_path
14
```

Figura 3. Redirecionamentos mais Seguros. Adaptado de [Bezold 2016]

5. Conclusão

Este estudo demonstrou o uso de controles de segurança para a proteção contra as vulnerabilidades citadas no relatório *OWASP Top 10 2013*. Entretanto, há outros tipos de controles de segurança, como, por exemplo, o registro de *logs* e detecção de instrusão, que não foram citados neste artigo, mas que podem ser conferidos em [Anton *et al. 2016*] e [Manico *et al.* 2016].

Trabalho correlacionado foi realizado por Silva e Garcia (2015), que apresenta um diagnóstico da eficácia dos mecanismos de proteção contra as vulnerabilidades mais frequentes entre o usuário e o navegador *web*. O trabalho ainda propõe o desenvolvimento de um ambiente controlado, simulando uma aplicação *web* vulnerável, a fim de realizar os testes com as ferramentas disponibilizadas pelos navegadores.

É necessário ressaltar a importância de outros fatores para a proteção mais ampla dos sistemas, como a análise e a modelagem de riscos [Howard 2006]. Esta etapa é para Howard e Lipner (2006) a mais importante no processo de inclusão da segurança em *software*, pois ela reduz significativamente os custos de desenvolvimento. Isto ocorre devido à capacidade de encontrar problemas ainda em fases iniciais do processo, reduzindo custo de manutenção.

Além das arquiteturas de segurança no ciclo de desenvolvimento de *software*, há ainda outros métodos de segurança que agregam uma maior proteção, como proposto no trabalho de [Wagner *et al.* 2015], que elaborou níveis de segurança por meio de modelos encontrados na literatura.

Como trabalho futuro, sugerimos o desenvolvimento de uma aplicação *web*, considerando os conceitos abordados neste artigo e utilizando os controles de segurança, de forma proativa, para evitar as vulnerabilidades. Para este processo, aconselhamos o emprego das diretrizes de segurança, como o GDO e o MSDL, a fim de promover uma proteção por todo o ciclo de vida da aplicação.

Defender proativamente os sistemas das ameaças tornou-se um requisito fundamental no processo de desenvolvimento de *software*. Com este trabalho foi possível ter uma visão dos controles essenciais na prevenção contra as principais ameaças de aplicações *web* e assim agregar uma camada de proteção a mais no processo de desenvolvimento de *software* orientado a segurança.

5. Referências

Anton, K. e Manico, J. e Bird, J. (2016). "OWASP Proactive Controls For Developers: 10 Critical Security Areas That Software Developers Must Be Aware Of", https://www.owasp.org/images/9/9b/OWASP_Top_10_Proactive_Controls_V2.pdf, Outubro.

Baan, S. e Chesney B. (2015). "OWASP Developer Guide 3.0", https://github.com/OWASP/DevGuide, Julho.

Barbosa, E. D. e Castro, R. de O. (2015). Desenvolvimento de Software Seguro: Conhecendo e Prevenindo Ataques Sql Injection e Cross-site Scripting (XSS). Tecnologias, Infraestrutura e Software, São Carlos, v. 4, n. 1, p. 31-40, jan-abr 2015.

Bezold, S. (2016). "Unvalidated Redirects and Forwards Cheat Sheet", https://www.owasp.org/index.php/Unvalidated_Redirects_and_Forwards_Cheat_Sheet, Setembro.

Carvalho, A. W. de. e Júnior, A. P. de C. (2014). "Google Hacking para ataques SQL Injection". II Escola Regional de Informática de Goiás, Goiânia, p. 65-77, nov 2014.

Howard, M. e Lipner, S. (2006). "The Security Development Lifecycle SDL: A Process for Developing Demonstrably More Secure Software", Microsoft Press, Redmond.

Machry. M (2013). “OWASP Top 10 2013: Os dez riscos de segurança mais críticos em aplicações *web*”, https://www.owasp.org/images/9/9c/OWASP_Top_10_2013_PT-BR.pdf, Setembro.

Manico, J. e Stock, A. e Cuthbert, D. (2016). “Application Security Verification Standard 3.0.1”, https://www.owasp.org/images/3/33/OWASP_Application_Security_Verification_Standard_3.0.1.pdf, Julho.

Mcgraw, G. (2006) “Software Security: building security in”. Pearson Education, Indiana.

Microsoft. (2010) “Implementação simplificada do Microsoft SDL”, https://www.microsoft.com/pt-BR/download/confirmation.aspx?id=12379, Fevereiro.

Silva, C. M. R. da. e Garcia, V. C. (2015). "Uma Avaliação de Proteção de Dados Sensíveis através do Navegador *web*". XV Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais, Florianópolis, p. 86-99, novembro 2015.

Sommerville, Ian. “Engenharia de Software”. Pearson, São Paulo, 2011.

Stallings, W. e Brown, L. (2014). “Segurança de Computadores: princípios e práticas”. Elsevier, Rio de Janeiro.

Stock, A. v. d. (2005). "OWASP Developer Guide 2.0", https://github.com/OWASP/DevGuide/tree/dc5a2977a4797d9b98486417a5527b9f15d8a251/DevGuide2.0.1, Junho.

Pressman, R. S. (2011). "Engenharia de Software: uma abordagem profissional". AMGH, Porto Alegre.

Wagner, R. e Machado, A. (2015). "Níveis de segurança para processos de desenvolvimento de software seguro". IX Simpósio de Informática, Rio Grande do Sul.

Wichers, D. e Petefish, P. e Sheridan, E. (2016). "Cross-Site Request Forgery (CSRF) Prevention Cheat Sheet", https://www.owasp.org/index.php/Cross-Site_Request_Forgery_(CSRF)_Prevention_Cheat_Sheet#General_Recommendations_For_Automated_CSRF_Defense, Setembro.

Winch. R. (2015). "Hello Spring Security Java Config", http://docs.spring.io/spring-security/site/docs/current/guides/html5//helloworld-javaconfig.html, Agosto.

Framework de Ferramentas FOSS para a Segurança de Servidores GNU/Linux em conformidade com os controles da norma ABNT NBR ISO/IEC 27002:2013

Maiky Nata M. Ribeiro[1], **Leonardo Afonso Amorim**[1], **Iwens Gervasio Sene Junior**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.001-970
Goiânia – GO – Brasil

maikyribeiro@inf.ufg.br[1], leonardoafonso@inf.ufg.br[1], iwens@inf.ufg.br[1]

Abstract. *The challenges in the Information Security area change very quickly and maintain servers safely requires constant updating of the methods and the software involved. Motivated by the lack of an updated and extensive FOSS (Free and Open-Source Software) Framework, this article aims to indicate current tools that were tested to suit GNU/Linux servers with standard ABNT NBR ISO/IEC 27002:2013. To perform the experiments with the technology proposed in this paper, a scenario composed of machines that represent servers within a LAN was set up and only the "Firewall Gateway" was exposed directly to the Internet. To test the security of servers configured according to the standard, intrusion attempts were made, both inside and outside the LAN (WAN), using machines with Kali Linux, a Linux distribution that specializes in penetration testing. As a result they are indicated FOSS tools that were able to mitigate possible risks to the proposed scenario.*
Keywords: FOSS Framework, ABNT NBR ISO/IEC 27002:2013, Hardening

Resumo. *Os desafios na área de Segurança da Informação mudam com muita rapidez e manter servidores com segurança requer atualização constante dos métodos e softwares envolvidos no processo. Motivado pela falta de um Framework FOSS (Free and Open-Source Software) atualizado e amplo, este artigo tem como propósito indicar ferramentas atuais e, principalmente, testadas para adequar servidores* GNU/Linux *com a norma ABNT NBR ISO/IEC 27002:2013. Para realizar os experimentos com as tecnologias propostas neste artigo, foi criado um cenário composto por máquinas que representam servidores dentro de uma LAN, tendo apenas o "Firewall Gateway" exposto diretamente na Internet. Para testar a segurança dos servidores configurados de acordo com a norma, foram realizados tentativas de invasão, tanto dentro da LAN quanto fora (WAN), usando máquinas com Kali Linux, uma distribuição Linux especializada em penetration testing. Como resultado são indicadas ferramentas FOSS que foram capazes de mitigar possíveis riscos ao cenário proposto.*
Palavras-chave: Framework FOSS, ABNT NBR ISO/IEC 27002:2013, Hardening

1. Introdução

Atualmente, disponibilizar um serviço de rede, como um servidor Web, servidor de e-mail, dentre outros, não é uma tarefa complicada. Existem muitas empresas como

Amazon Web Services e *Digital Ocean* que trabalham com *VPS Hosting*[1] e com poucos cliques é possível colocar no ar diversos serviços de redes importantes por um preço acessível [Santos 2015]. No entanto, trabalhar com serviços de rede com configurações padrões pode colocar em risco a continuidade de negócio de uma empresa. Logo, manter uma rede de computadores corporativa é um grande desafio devido a inúmeras variáveis que devem ser levadas em consideração quando se tem ativos valiosos de uma empresa: as informações do negócio. Dentre estas variáveis é possível destacar algumas como, por exemplo: organização dos serviços de rede entre os servidores disponíveis, definir política de usuários e senhas, rotinas de *backup*, administração dos registros de atividades (*logs*), gerenciamento de sistemas de detecção de intrusão, análise de vulnerabilidades dos sistemas que compõe a rede, entre outras [Sandro Melo 2006].

O sistema *GNU/Linux* tem sido preferência para serviços de rede devido seu ótimo desempenho, robustez, segurança, grande quantidade de documentação e custo-benefício. No entanto, por mais seguro que o *GNU/Linux* seja, é necessário fortalecer o sistema com práticas de segurança adequadas que sejam guiadas por uma norma de segurança madura. É possível encontrar com certa facilidade diversos trabalhos que apresentam ferramentas para implementar as políticas de segurança. Em [Barbosa 2012], é apresentado de maneira teórica um ambiente organizacional, políticas de segurança, vulnerabilidades, engenharia social e *hardening*[2] de servidores, porém o autor foca em apenas 5 itens de controles da norma 27002:2005. Em [Junior 2008], o foco é a norma ISO 27001, que é a base para a norma 27002, porém voltada para um processo de implementação de um Sistema de Gestão da Segurança da Informação (SGSI), e contempla 33 controles da norma 27001 com sua respectiva sugestão para conformidades. Em [Titon 2013], falta um embasamento em controles e boas práticas de uma norma como a 27002 ou a 27001, ou seja, apenas busca aplicar as melhores técnicas de *hardening* em servidores, tanto *GNU/Linux* como *Windows*. *Hardening* é apenas um dos itens da norma ABNT NBR ISO/IEC 27002.

Os estudos citados permitem estabelecer a proposta de um Framework atualizado, expandindo os limites dos mesmos, a fim de demonstrar maior contribuição para a comunidade. Os desafios na área de Segurança da Informação mudam com muita rapidez e manter servidores com segurança requer atualização constante. Atualmente, a norma ABNT NBR ISO/IEC 27002 está na versão 2013. Baseado nestes desafios este artigo tem como propósito indicar ferramentas FOSS atuais e testadas para adequar os sistemas com a norma ABNT NBR ISO/IEC 27002:2013 e orientar analistas e administradores de sistemas de maneira prática e eficiente. Na Seção 2 é apresentada a norma ABNT NBR ISO/IEC 27002:2013. Na Seção 3 é apresentada a topologia da rede para realização dos testes a fim de facilitar o entendimento da aplicação das ferramentas que são apresentadas neste artigo. Na Seção 4, experimentos e ferramentas são explicadas de acordo com a norma. Na Seção 5 é feita uma discussão sobre os experimentos realizados com as ferramentas sugeridas e na Seção 6 e 7 é feita a conclusão e possíveis trabalhos futuros respectivamente.

[1]VPS (Virtual Private Server) é uma máquina virtual vendida como serviço por uma empresa de hospedagem.

[2]Hardening significa endurecer. Neste contexto endurecer as políticas de segurança dos servidores.

2. A Norma ABNT NBR ISO/IEC 27002:2013

A norma ABNT NBR ISO/IEC 27002:2013 é um código de prática para controles de segurança da informação e fornece diretrizes para práticas de gestão de segurança da informação e normas de segurança da informação para as organizações, incluindo a seleção, implementação e o gerenciamento de controles, levando em consideração os ambientes de risco da segurança da informação da organização [ABNT:27002 2013].

Dependendo das circunstâncias, os controles de segurança da informação de quaisquer seções podem ser importantes; assim, convém que cada organização implemente esta norma identificando quais controles são aplicáveis, quão importantes eles são e qual a aplicação para os processos individuais do negócio [ABNT:27002 2013].

A norma é projetada para ser usada por organizações que pretendam: (i) selecionar controles dentro do processo de implementação de um sistema de gestão da segurança da informação baseado na ABNT NBR ISO/IEC 27001; (ii) implementar controles de segurança da informação comumente aceitos; (iii) desenvolver seus próprios princípios de gestão da segurança da informação [ABNT:27002 2013]. Além disso, ela é projetada para as organizações usarem como referência na seleção de controles dentro do processo de implementação de um Sistema de Gestão da Segurança da Informação (SGSI) baseado na ABNT NBR ISO/IEC 27001 ou como um documento de orientação para as organizações implementarem controles de segurança da informação comumente aceitos [ABNT:27002 2013].

A norma contém 14 seções de controles de segurança da informação, 35 objetivos de controles e 114 controles. Cada seção principal contém: (i) um objetivo de controle declarando o que se espera ser alcançado; (ii) um ou mais controles que podem ser aplicados para se alcançar o objetivo do controle [ABNT:27002 2013].

Considerando este escopo da norma, e analisando os controles indicados por ela, este Framework foi criado dando conformidade ao sistema operacional GNU/Linux com os controles possíveis de serem implementados no sistema, juntamente com as ferramentas FOSS indicadas.

3. Topologia da rede para realização dos testes

O cenário é composto por máquinas virtuais com sistema operacional *Debian GNU/Linux 8.0* e foi construído a fim de facilitar o entendimento das aplicações das ferramentas FOSS em um ambiente real. Com isso, os requisitos da norma ABNT NBR ISO/IEC 27002:2013 neste cenário serão validados. As máquinas virtuais são gerenciadas pelo software de virtualização VirtualBox [Oracle 2016b]. Foi criada uma máquina "Firewall Gateway" com o objetivo de minimizar as tentativas de ataques que a rede local recebe, impedindo possíveis invasões e levantamento de informações. Dessa forma a rede foi segregada e as máquinas "DMZ1" e "DMZ2" conforme Figura 1 estão em uma zona desmilitarizada a fim de dar conformidade ao controle 13.1.3 da norma, que descreve diretrizes a serem criadas para que sistemas de informação sejam segregados em redes.

Por meio das máquinas "Atacante Externo" e "Atacante Interno" conforme mostradas na Figura 1 instaladas com a distribuição *Kali Linux* [OffensiveSecurity 2016], foram executados testes de exploração de vulnerabilidades (*penetration testing*) para avaliar as ferramentas FOSS configuradas nas máquinas "Firewall Gateway", "DMZ1"

e "DMZ2" que são representadas na Figura 1. Assim, dando conformidade com o controle 12.6.1 da norma, que diz respeito à obtenção de informações sobre vulnerabilidades técnicas dos sistemas de informação em uso, expondo a organização à estas vulnerabilidades, e logo depois avaliando e tomando medidas apropriadas para lidar com os riscos associados.

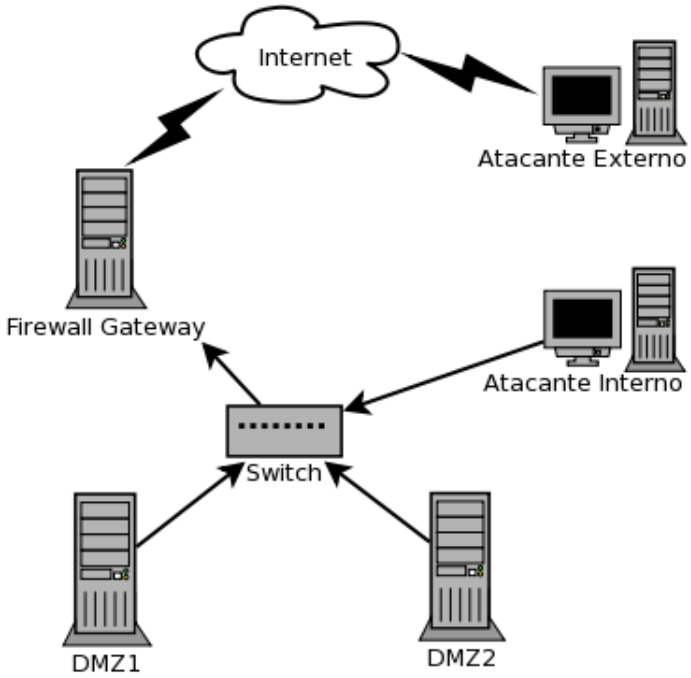


Figura 1. Topologia da rede para realização dos testes para aplicação dos controles da norma ABNT NBR ISO/IEC 27002:2013

4. Experimentos e Ferramentas

De acordo com o trabalho realizado no cenário exposto, nesta Seção são apresentados as ferramentas propostas para a conformidade com a norma, de acordo com seus controles selecionados, bem como os experimentos realizados para alcançar a validação das ferramentas indicadas.

4.1. Trabalho remoto e proteção das transações nos aplicativos de serviços

Medidas devem ser implementadas para proteger as informações acessadas, processadas ou armazenadas em locais de trabalho remoto. Dentre várias providências a serem tomadas para conseguir o objetivo desse controle 6.2.2 descrito na norma, no caso do cenário abordado, levou-se em maior consideração, uma rede de trabalho remoto com informações que devem ser protegidas durante a comunicação. A ferramenta indicada é o *OpenVPN* [OpenVPN 2016]. A VPN (*Virtual Private Network*) é uma rede de comunicação particular, geralmente utilizada em canais de comunicação inseguros, como a WAN (Internet) ou mesmo a LAN (Rede Local).

O que torna esta rede de comunicação particular é o fato das ferramentas de VPN empregarem métodos e protocolos de criptografia, criando um "túnel" para prover acesso seguro às partes da rede separadas. Para tornar ainda mais seguro, é indicado o uso

do *OpenSSL* [OpenSSL 2016] para gerar certificados e chaves que serão utilizados nas configurações do OpenVPN.

Foi realizado o seguinte experimento no cenário: Ao ser feita a instalação do servidor *OpenVPN* na máquina "Firewall Gateway" foi gerado um certificado para o servidor e um certificado para a máquina "DMZ1"', onde foi configurado o cliente *OpenVPN*. Dessa forma, foram feitos testes de troca de informações através do *OpenSSH* [OpenBSD 2016]. A máquina "Atacante Interno" estava monitorando os pacotes da rede com a ferramenta *Wireshark* [Wireshark 2016], e não foi possível decifrar o conteúdo das informações devido a criptografia empregada nos pares de chave, que protegem a conexão. Com isso, também foi dado conformidade ao controle 14.1.3 da norma, que trata da prevenção de transmissões incompletas, erros de roteamento, alteração não autorizada da mensagem, divulgação não autorizada, duplicação ou reapresentação da mensagem não autorizada.

4.2. Inventário dos ativos

Para atingir o objetivo do controle 8.1.1, a norma estabelece que os ativos associados com informação e com os recursos de processamento da informação sejam identificados e um inventário destes ativos seja estruturado e mantido. Para esse controle, indica-se a ferramenta *OCS Inventory NG* [OCS-Inventory-NG 2016], um software livre que permite aos administradores ter um inventário de seus ativos de TI automaticamente. O *OCS-NG* recolhe informações sobre o hardware e software de ativos que executam o programa cliente *OCS (OCS Inventory Agent)*. O seguinte experimento foi realizado: Na máquina "Firewall Gateway" do cenário, já com os pré-requisitos de softwares instalados e configurações necessárias feitas, foi instalado o *OCS Inventory NG Server*, ativando certificados SSL no servidor web Apache e gerando as chaves para que a comunicação entre os agentes e o servidor fosse segura.

Após isso o agente *OCS* foi instalado nas máquinas "DMZ1" e "DMZ2"', a partir daí, na interface web do servidor *OCS*, foi visto todos os dados das duas máquinas, como versões de hardware, sistema operacional, softwares e pacotes instalados, endereço IP e demais dados de rede das máquinas. Todo esse inventário pode ser mantido em um banco de dados do *OCS Inventory NG Server*, assim como todos os registros inerentes a esses ativos.

4.3. Classificação da Informação e política de controle de acesso

Na norma os controles 8.2.1 e 9.1.1 envolvem um trabalho com sensibilidade e criticidade das informações, para evitar modificação e divulgação não autorizada, dessa forma é preciso trabalhar com permissões de sistema e uma política de controle de acesso a ser estabelecida, documentada, analisada e mantida. Em ambientes mais complexos, o controle de acesso das informações da empresa, principalmente o acesso a diretórios e compartilhamentos de arquivos se torna bastante crítico. Nesse contexto, é indicada a instalação do pacote ACL (*Access Control Lists*) que é uma ferramenta de apoio para essa análise, aplicação de controles de permissões e acessos efetivos.

Um caso de teste foi feito baseado no seguinte exemplo: Imagine um departamento de contabilidade e um departamento financeiro dentro da empresa. O funcionário João, da contabilidade, precisa acessar alguns arquivos dentro de um

subdiretório do diretório do departamento financeiro, porém ele não pode ter direito de alterar ou apagar os arquivos que estão dentro do diretório financeiro. Para que isso seja possível, na máquina "DMZ1" foi criado um diretório chamado "documentos", definindo que essa determinada partição trabalha com ACLs de *filesystem*. Logo após criou-se dois grupos, um chamado financeiro e outro contabilidade, em seguida também foi criado os usuários pertencentes a cada grupo. Dentro do diretório "documentos" foram criados os diretórios "financeiro" e "contabilidade", e seus grupos associados aos respectivos diretórios. Fazendo as configurações necessárias com a ferramenta ACL, foi verificado que o usuário João conseguiu acesso ao diretório financeiro, porém não conseguiu alterar ou apagar os documentos, somente visualizá-los. Com esse resultado também dá-se conformidade ao controle 9.1.2 da norma que trata de conceder acesso às redes somente para os usuários especificamente autorizados a usar.

4.4. Gerenciamento de usuários, privilégios e autenticação

Para conseguir esse controle, foram utilizadas as próprias ferramentas e comandos nativos do *GNU/Linux*, como adduser, passwd, usermod, groupadd, gpasswd, chown, chmod, chgrp, chsh e umask, a fim de implementar um processo de registro e cancelamento de usuário, atribuição de direitos de acesso e também o controle para conceder ou revogar direitos e privilégios de determinados usuários, conforme os controles 9.2.1 e 9.2.2, seguindo também a orientação do item 9.2.5 da norma. Foi implementado também datas para a expiração de contas de usuários, utilizando o comando chage, dando conformidade com o item 9.2.6 da norma que trata da retirada de acessos de usuários logo após o encerramento de suas atividades, dessa forma feito de modo automático por meio deste comando. Também foi bloqueado o login do usuário root nos terminais de texto das máquinas "Firewall Gateway", "DMZ1" e "DMZ2", conforme controle 9.2.3 da norma, que diz respeito ao gerenciamento de direitos de acesso privilegiados, pois deixar o login como root habilitado para os terminais de texto é bastante inseguro, e quando for necessário a execução de uma tarefa administrativa o usuário comum poderá virar root, utilizando-se do comando su ou o sudo.

Os usuários da rede acessam a Internet por meio de um servidor *Proxy Squid* [Squid 2016]. O *Proxy Squid* é importante pois exige que o usuário tenha uma conta para acessar o conteúdo na Internet. Com isso, todo o acesso será registrado por meio de *logs*. Os registros permitem ao administrador do sistema fazer auditoria nos acessos dos usuários da rede. O proxy foi implementado na máquina "Firewall Gateway".

Para centralizar as contas de usuários de todos os serviços de rede importantes (Servidor de e-mail, web, ftp, proxy, VPN etc) foi usado o *OpenLDAP* [OpenLDAP 2016], uma implementação livre do protocolo LDAP. Informações de usuários e, principalmente, contas de usuários, quando centralizadas facilitam procedimentos importantes como *backup* e replicação da base de dados para outro servidor para que exista redundância do serviço. O *OpenLDAP* foi implementado nas máquinas "DMZ1" e "DMZ2" sendo que a primeira funciona como servidor principal e a segunda como secundário. A replicação da base de dados entre as máquinas "DMZ1" e "DMZ2" foi feita com o *Syncrepl*. Para garantir que o serviço *OpenLDAP* fique disponível pelo maior tempo possível foi configurado na máquina "Firewall Gateway" um balanceador de carga com o *IPVS (IP Virtual Server)* [IPVS 2016] e para ter alta disponibilidade foi usado o *Keepalived* [Keepalived 2016].

4.5. Sistema de gerenciamento de senha

Utilizando o PAM (*Pluggable Authentication Modules for Linux*) nas máquinas "Firewall Gateway", "DMZ1" e "DMZ2", primeiramente descobriu-se quais aplicações que estavam instaladas tinham suporte ao PAM nas máquinas citadas, após esse resultado é possível gerenciar os programas críticos para o sistema, conforme o item 9.4.4 da norma. Foram feitos testes de limite de login de alguns usuários, bloqueio de horário ou dia em que o usuário não pode logar, limitação de logins consecutivos, dando conformidade ao item 9.4.2 da norma. E também foi feita a limitação de usuários que podem usar o comando su ou sudo para instalação ou modificação de pacotes, dando conformidade aos controles 12.5.1 e 12.6.2 da norma.

Além disso um dos testes mais importantes foi o trabalho com senhas mais seguras, tratado no controle 9.4.3 da norma, quanto ao gerenciamento de senhas e a diretriz para que seja assegurado senhas de qualidade, o qual o módulo do PAM pode gerenciar fazendo-se poucas configurações. Com isso, foram feitos testes de força bruta, usando o método de dicionário, com a ferramenta *John the Ripper* [Openwall 2016] para avaliar se as senhas de usuários estão vulneráveis (senhas baseadas em sequências numéricas e baseadas em palavras de dicionário) contra os usuários que criaram senhas mais seguras gerenciado conforme os requisitos que foi implementado o módulo PAM (senhas mais complexas e com caracteres especiais), pois conforme o item 9.2.4 da norma, a concessão de informação de autenticação secreta deve ser controlada por um processo formal. No resultado do experimento foi verificado que, por meio de força bruta (baseado em dicionário), as senhas fracas foram "quebradas" em um pequeno período de tempo pela ferramenta *John the Ripper*, porém tornou-se muito mais difícil a quebra de senhas mais fortes e complexas.

4.6. Política para o uso de controles criptográficos e gerenciamento de chaves

Para a implementação de controles criptográficos e gerenciamento de chaves, conforme indicação dos itens 10.1.1 e 10.1.2 da norma, foi utilizada o *GnuPG* [GnuPG 2016], que permite criptografar e assinar seus dados e informações de comunicação e possui um sistema de gerenciamento de chaves, além de módulos de acesso para todos os tipos de diretórios de chaves públicas. O *GnuPG*, também conhecido como *GPG*, também possui características para fácil integração com outros aplicativos. Foi realizado o seguinte experimento: Na máquina "DMZ1" foi gerada uma chave para um usuário padrão, também na "DMZ2" o mesmo foi feito para outro usuário, em seguida a chave pública do usuário da "DMZ1" foi exportada para um arquivo, na "DMZ2" foi feito o mesmo procedimento. Em seguida as chaves públicas desses usuários foram trocadas entre eles. Logo após o usuário da "DMZ1" utilizou o *GPG* para gerar um arquivo encriptado com sua chave, e o enviou para o usuário da "DMZ2". Após o recebimento do arquivo na "DMZ2", foi feito um teste de leitura do conteúdo antes de decriptá-lo, nesse momento observou-se que não era possível decifrar o conteúdo devido a criptografia empregada na geração do arquivo. No entanto utilizando a chave pública do usuário da "DMZ1", importada anteriormente, o usuário da "DMZ2" conseguiu decriptar o arquivo com o *GPG* e ver seu conteúdo na forma original.

4.7. Segurança física

Referente ao acesso físico, a norma trata na seção 11 como prevenir o acesso físico não autorizado, danos e interferências com os recursos de processamento das informações

da organização. Seguindo esse controle, a nível do sistema operacional *GNU/Linux*, a segurança do *GRUB* [FSF 2010], o gerenciador de inicialização padrão do *GNU/Linux*, foi melhorada, pois se algum intruso acessar fisicamente o servidor poderá quebrar a senha do administrador, para causar danos ao sistema, por meio das vulnerabilidades encontradas nas instalações padrões do *GRUB*. Para corrigir isso, o seguinte experimento foi realizado: O *GRUB* foi reinstalado e uma camada de criptografia foi adicionada para a proteção da senha a fim de evitar a edição de suas linhas e parâmetros quando ocorre a inicialização do sistema. Com isso foi observado a impossibilidade de editar os parâmetros do gerenciador de inicialziação *GRUB* sem que a senha fosse digitada.

4.8. Controles contra códigos maliciosos

Como controle contra códigos maliciosos, foi utilizada a ferramenta *OSSEC* [OSSEC 2016], que é um *HIDS (Host-based Intrusion Detection System)*, que tem muitas funcionalidades, dentre elas a detecção de *rootkits*, que são códigos maliciosos que exploram falhas do sistema operacional para obter acesso como administrador. Com essa configuração habilitada no *OSSEC*, um controle de detecção e prevenção contra códigos maliciosos foi implementado conforme o item 12.2.1 da norma. O *OSSEC* foi instalado na máquina "DMZ2" e um experimento foi feito para verificar a capacidade em detectar um *rootkit*. A máquina "Atacante Externo" enviou um e-mail contendo no anexo um *rootkit*, quando o servidor de e-mail instalado na "DMZ2" recebeu essa mensagem, automaticamente o *OSSEC* fez a detecção do pacote malicioso e alertou o administrador da rede sobre essa ameaça.

Como demonstrado, o *OSSEC* tem a capacidade de trabalhar localmente ou em rede como cliente e servidor, e uma das grandes vantagens dessa ferramenta é o *active-response*, ou seja, para determinados tipos de ataques ele pode tomar algumas medidas como bloquear o endereço IP que está atacando por um determinado tempo e mandar um e-mail alertando sobre o ocorrido. Com essa configuração feita, dá-se conformidade ao controle 16.1.1, que diz respeito a procedimentos de gestão estabelecidos para assegurar respostas rápidas e efetivas a incidentes de segurança da informação. Para validar o trabalho do *OSSEC* em rede, o seguinte experimento foi feito: Com o *OSSEC* instalado na máquina "Firewall Gateway", foi verificado como o seu *active-response* trabalha. O "Atacante Externo" fez tentativas conexão *telnet* no servidor de e-mail disponibilizado na Internet instalado na máquina "DMZ2", essas tentativas de conexão, foram interpretadas pelo *OSSEC* como um ataque, e o endereço IP da máquina "Atacante Externo" foi bloqueado com uma regra criada nas configurações do *iptables* da máquina "Firewall Gateway", que gerencia o tráfego de internet para a máquina "DMZ2".

Em ambos os testes o *OSSEC* enviou um e-mail ao administrador alertando sobre as ameaças, dando conformidade ao controle 16.1.2, que diz respeito a notificação de eventos de segurança da informação.

4.9. Resposta aos incidentes de segurança da informação

Para a implementação do controle 16.1.5, que descreve a resposta aos incidentes de segurança reportando os mesmos para os administradores, na máquina "Firewall Gateway", foi utilizada a ferramenta *Snort* [Cisco 2016b], que realiza a tarefa de detectar e registrar possíveis ataques à rede. O *Snort* gera *logs* relativos a possíveis de ataques, que por padrão são em formato de texto. Para facilitar a visualização deles, uma configuração

foi feita para que os *logs* pudessem ser gravados em uma base de dados *MySQL* [Oracle 2016a]. Foi executado o seguinte experimento para verificar o potencial de detecção de vulnerabilidades do *Snort*: A máquina "Atacante Externo" fez uma varredura com a ferramenta *Nmap* [LLC 2016] buscando pelos serviços de rede instalados e portas abertas na máquina "Firewall Gateway", em tempo real o *Snort* gerou registros para essa varredura realizada e tudo foi registrado no banco de dados, para que posteriormente o administrador pudesse analisar e tomar as devidas decisões quanto aos incidentes.

4.10. Cópias de segurança das informações

Para o controle 12.3.1 da norma, que trata de cópias de segurança das informações, para que sejam efetuadas e testadas regularmente, é indicado a ferramenta *Bacula* [Bacula 2016], que permite o administrador de sistemas gerenciar *backups*, recuperação e verificação de dados. Para o teste o *Bacula Server* foi instalado na máquina "DMZ2" e o *Bacula Client* na máquina "DMZ1". O cliente foi adicionado no servidor *Bacula* e criado uma tarefa de *backup* padrão para ser executada semanalmente. A tarefa foi verificada após sua execução e observou-se que o *backup* foi realizado com sucesso. Logo após, foi realizado um teste de restauração do *backup* do cliente, e o mesmo foi efetuado com sucesso.

4.11. Registros de eventos

A norma recomenda nos controles 12.4.1, 12.4.2 e 12.4.3 características de um sistema de registro de *logs* de eventos das atividades do usuário, exceções, falhas e eventos de segurança da informação, além de serem mantidos, protegidos e analisados criticamente. Para dar conformidade a esses controles foi instalada na máquina DMZ2 a ferramenta *Syslog-NG* [BalaBit 2016], para ser utilizada como servidor de *logs*, armazenando registros de atividades dos outros computadores na rede como, por exemplo, das máquinas "Firewall Gateway" e "DMZ1", onde o *Syslog-NG* foi configurado para enviar os *logs* para o servidor "DMZ2". Por motivo de segurança, é necessário que os *logs* não fiquem apenas nas próprias máquinas, pois eles podem ser alterados por um atacante. Portanto, é importante manter uma cópia em outro local. O seguinte experimento foi realizado, na máquina "Firewall Gateway" foi executado um evento de teste para gerar um registro de log, o mesmo foi feito na máquina "DMZ1". Foi observado que os *logs* foram enviados para o servidor *Syslog-NG* na máquina "DMZ2" com sucesso.

4.12. Sincronização dos relógios

A sincronização de relógios também é uma das diretrizes cobradas pela norma no controle 12.4.4 e descreve que sejam sincronizados com uma única fonte de tempo precisa. Para isso foi utilizado o *ntpd* [NTP 2016], um *daemon*[3] para o sistema operacional que estabelece e mantém a hora do sistema em dia e sincronizado com servidores de tempo na internet, e é uma implementação completa para uso do *Network Time Protocol (NTP)* [NTP 2016]. Para que os relógios fossem sincronizados, o *ntpd* foi instalado nas máquinas "Firewall Gateway", "DMZ1" e "DMZ2", e foram configurados os mesmos servidores *NTP* para todas estas máquinas. Dessa forma foi observado um sincronismo de data e horário equivalente para todas as máquinas citadas.

[3]Um processo servidor que executa em segundo plano.

4.13. Controles de redes

Um *Firewall* faz o filtro de pacotes que passam na rede, para configurá-lo é necessário o conhecimento sobre a estrutura da rede em questão e dos diferentes protocolos envolvidos na comunicação, isto é, dos serviços que a rede usa para que eles não percam a comunicação. O objetivo em ter uma máquina fazendo o papel de "Firewall Gateway" na rede é minimizar as tentativas de ataques que ela recebe, tentando impedir possíveis invasões e levantamento de informações, implementando isso, atende-se ao controle 13.1.1 da norma que descreve que convém que as redes sejam gerenciadas e controladas para proteger as informações nos sistemas e aplicações. Para isso foi usada a ferramenta nativa do GNU/Linux, o *iptables* [Netfilter 2015], que com suas tabelas de regras pode filtrar, controlar, permitir e negar os pacotes da rede em que está instalado.

Antes da implementação do *iptables* na máquina "Firewall Gateway", a máquina "Atacante Externo" fez varreduras usando a ferramenta *OpenVAS* [OpenVAS 2016], e encontrou diversas vulnerabilidades, além de descobrir os serviços ativos e obter informações sobre os softwares instalados na máquina. Ao total foram 11 resultados de alto risco, 22 de médio risco e 3 de baixo risco.

Para corrigir isso, o seguinte experimento foi realizado: Na máquina "Firewall Gateway" foi implementado um script de *firewall* para o *iptables*, para inicialização durante a inicialização da máquina. Este script tem por objetivo negar todo tráfego para as regras da tabela *filter* do *iptables*, e liberar somente os serviços previamente autorizados pela política da rede, o que são chamadas de exceções. Todo o tráfego de pacotes que as exceções não cobrirem é filtrado e negado ("drop") automaticamente. Dessa forma, após a aplicação do script ser realizada, o *firewall* foi inicializado e foi realizado uma nova varredura pela máquina "Atacante Externo" utilizando a ferramenta *OpenVAS*. Foi observado que os números de vulnerabilidades caíram drasticamente, restando apenas 6 de risco nível médio e 2 de nível baixo. Com isso os resultados obtidos por essas análises realizadas por *pentest* podem reduzir a probabilidade ou impacto de incidentes futuros, dando conformidade ao controle 16.1.6 da norma.

4.14. Mensagens eletrônicas

Para adequar o gerenciamento de mensagens eletrônicas ao controle 13.2.3 e prover proteção às informações que trafegam em mensagens eletrônicas, foi utilizada a ferramenta *OpenSSL*, instalada em conjunto com o nosso servidor de e-mail *Postfix* [Venema 2016], para adicionar uma camada *SSL*, gerando certificados entre os usuários. Depois disso, foi adicionada outra camada de antivírus e antispam com as ferramentas *Clamav* [Cisco 2016a] e *Apache SpamAssassin* [Apache 2015]. No experimento realizado a máquina "Atacante Externo" enviou um e-mail para um usuário da máquina "DMZ1", com um domínio não registrado e não confiável, e observou-se que a ferramenta *Apache SpamAssassin* fez o filtro desse e-mail colocando-o na pasta "spam" da caixa eletrônica do usuário.

4.15. Gerenciamento de capacidade e segurança dos serviços de rede

A utilização dos recursos deve ser monitorada e ajustada para verificar as necessidades de capacidade e garantir o desempenho dos sistemas, conforme descrito no controle 12.1.3 da norma. Para atender esse controle foi utilizado a ferramenta *Nagios* [Nagios 2016], que

é uma poderosa ferramenta de monitoramento de redes e infraestrutura, sendo capaz de monitorar servidores, *switches*, aplicações, e atendendo aos requisitos do controle 13.1.2, também é uma ferramenta de gerenciamento dos serviços de uma rede. Para observar o *Nagios* em ação, o *Nagios Server* foi instalado na máquina "Firewall Gateway", para que fosse feito o monitoramento das máquinas "DMZ1" e "DMZ2", onde foram instalados o *Nagios Client*. Ao abrir o gerenciador web do *Nagios Server*, foram vistas todas as informações relacionadas às máquinas "DMZ1" e "DMZ2", isso quanto a monitoramento de servidor. Mas como monitoramento de serviço, é preciso adicionar o serviço a ser monitorado na rede. Dessa forma, foi adicionado o serviço *SSH* a ser monitorado, e logo depois foi feita uma conexão remota *SSH* da máquina "DMZ1" para a "DMZ2". Ao atualizar a página de gerenciamento web do *Nagios*, foi visualizado o serviço de rede *SSH* executando e quais máquinas o estavam utilizando.

4.16. Disponibilidade dos recursos de processamento da informação

Alta disponibilidade está relacionada com redundância e é recomendada para que possa suprir os requisitos do negócio de uma empresa, conforme é descrito no controle 17.2.1 da norma. Ao elaborar o cenário proposto por esse Framework, foram utilizadas máquinas virtuais com o software de virtualização VirtualBox para efetuar os testes, e caso ocorresse algum problema com alguma máquina, rapidamente seria possível voltar a um estado anterior utilizando um *snapshot* da máquina gerado antes do incidente.

Por isso foi feito o seguinte experimento: No gerenciador do VirtualBox foi feita uma cópia de segurança da máquina "DMZ2" gerando um *snapshot*, e se a partir desse momento a máquina tivesse algum problema ou incidente, o seu "snapshot" seria rapidamente restaurado. Simulando que a máquina teve um problema, logo conseguiu-se restaurá-la com sucesso. Para grandes empresas a mesma lógica é necessária, porém com um nível mais alto de virtualização profissional. Por isso o mesmo experimento foi feito utilizando uma das melhores ferramentas de virtualização do mundo, o *XenServer* [Project 2013]. Para isso foi instalada uma nova máquina, com o *XenServer* configurado, e realizado o mesmo experimento feito com o *VirtualBox*, e foi observado que a máquina foi restaurada com sucesso ao ponto anterior de pleno funcionamento. Assim, esta ferramenta consegue suprir e dar uma alta disponibilidade para a empresa que a utiliza.

5. Discussão

Dada uma vasta quantidade de ferramentas FOSS, escolher uma que seja madura e atenda aos requisitos da norma ABNT NBR ISO/IEC 27002:2013, demanda pesquisa e testes em cenários complexos. Portanto, os testes apresentados por esse Framework possibilitaram elucidar a maturidade das ferramentas, além da capacidade de fortalecer a segurança do sistema operacional *GNU/Linux*. Uma das dificuldades encontradas foi que nem todos os controles da norma puderam ser cobertos devido a impossibilidade de aplicá-los ao *GNU/Linux* e algumas ferramentas apresentadas cobriram apenas parcialmente alguns controles citados devido a ampla generalização descrita nos objetivos dos controles da norma. Outra observação feita é que uma ferramenta acaba dando conformidade a vários controles da norma simultaneamente, o que demonstra o grande poder de determinadas ferramentas FOSS indicadas.

Além disso, houve dificuldade em dosar os seguintes fatores importantes nesse processo: segurança, risco e flexibilidade. Pois uma configuração padrão de um

servidor pode trazer grande flexibilidade, porém a segurança ficará baixa e os riscos aumentarão. Ao atingir um alto nível de segurança para o sistema menor será o risco e, consequentemente, a flexibilidade de administração do sistema cai. Para facilitar esse entendimento, veja a Figura 2:

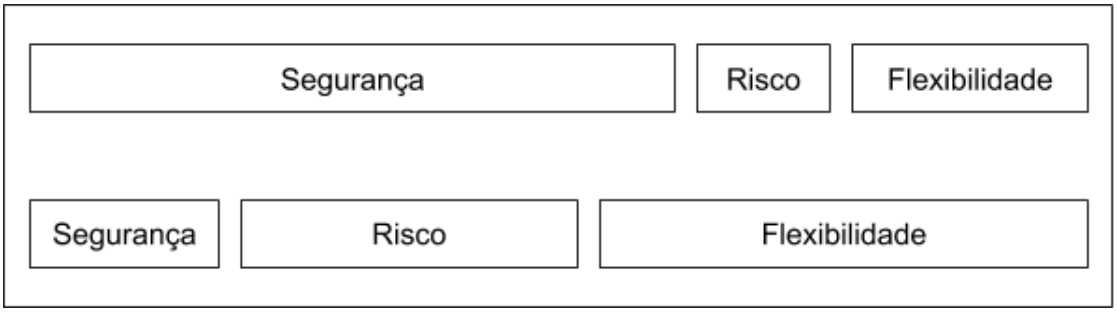


Figura 2. Análise de variáveis importantes para o fortalecimento da Segurança de um servidor [Sandro Melo 2006]

Levando isso em consideração, o que cada empresa e administrador de sistemas deve avaliar para que possa decidir se seu ambiente operacional poderá utilizar determinadas ferramentas ou determinados controles citados é essa possibilidade equilibrada de implantá-los e proteger a organização das ameaças e vulnerabilidades e, assim, reduzir o impacto de incidentes de segurança aos seus ativos.

6. Conclusão

Feito a análise da norma aplicando os seus controles e os testes com as ferramentas FOSS, concluiu-se que esse Framework possibilita ao administrador de sistemas um meio de dar conformidade ao *GNU/Linux*, baseado em uma norma que serve como código de prática e oferece diretrizes em uma ampla gama de controles de segurança da informação, que são normalmente aplicados em muitas organizações diferentes.

7. Trabalhos Futuros

Como trabalho futuro é recomendável que esse Framework seja atualizado e testado constantemente, pois várias vulnerabilidades são encontradas e novos programas e mecanismos de segurança surgem com frequência.

Referências

ABNT:27002 (2013). *ABNT - Tecnologia da Informação-Técnicas de Segurança - Código de Prática para controles de segurança da informação*. ABNT, http://www.abntcatalogo.com.br/norma.aspx?ID=306582, 1 edition.

Apache (2015). Apache spamassassin. `http://spamassassin.apache.org/`, Último acesso em Outubro de 2016.

Bacula (2016). Bacula. `http://blog.bacula.org/`, Último acesso em Outubro de 2016.

BalaBit (2016). Syslog-ng. `https://syslog-ng.org/`, Último acesso em Outubro de 2016.

Barbosa, F. S. (2012). Fundamentos em segurança e hardening em servidores linux baseado na norma iso 27002. *Anais Eletrônicos EnuComp*, 6(2).

Cisco (2016a). Clamav. https://www.clamav.net/, Último acesso em Outubro de 2016.

Cisco (2016b). Snort. https://www.snort.org/, Último acesso em Outubro de 2016.

FSF (2010). Grub. https://www.gnu.org/software/grub/, Último acesso em Outubro de 2016.

GnuPG (2016). Gnupg. https://www.gnupg.org/, Último acesso em Outubro de 2016.

IPVS (2016). Ip virtual server. http://www.linuxvirtualserver.org/software/ipvs.html, Último acesso em Novembro de 2016.

Junior, D. M. M. (2008). Proposta de hardening em conformidade com a iso 27001 para um firewall em linux com balanceamento de carga. Master's thesis, UFLA, http://repositorio.ufla.br/handle/1/5605.

Keepalived (2016). Keepalived. http://www.keepalived.org/, Último acesso em Novembro de 2016.

LLC, I. (2016). Nmap. https://nmap.org/, Último acesso em Outubro de 2016.

Nagios (2016). Nagios. https://www.nagios.org/, Último acesso em Outubro de 2016.

Netfilter (2015). iptables. http://ipset.netfilter.org/iptables.man.html, Último acesso em Outubro de 2016.

NTP (2016). ntpd. http://doc.ntp.org/4.1.0/ntpd.htm, Último acesso em Outubro de 2016.

OCS-Inventory-NG (2016). Ocs inventory ng. http://www.ocsinventory-ng.org/, Último acesso em Outubro de 2016.

OffensiveSecurity (2016). Kali linux. https://www.kali.org/, Último acesso em Outubro de 2016.

OpenBSD (2016). Openssh. http://www.openssh.com/, Último acesso em Outubro de 2016.

OpenLDAP (2016). Openldap. http://www.openldap.org/, Último acesso em Novembro de 2016.

OpenSSL (2016). Openssl. https://www.openssl.org/, Último acesso em Outubro de 2016.

OpenVAS (2016). Openvas. http://www.openvas.org/, Último acesso em Outubro de 2016.

OpenVPN (2016). Openvpn. http://community.openvpn.net/openvpn, Último acesso em Outubro de 2016.

Openwall (2016). John the ripper. http://www.openwall.com/john/, Último acesso em Outubro de 2016.

Oracle (2016a). Snort. https://www.mysql.com/, Último acesso em Outubro de 2016.

Oracle (2016b). Virtualbox. https://www.virtualbox.org/, Último acesso em Outubro de 2016.

OSSEC (2016). Ossec. http://ossec.github.io/index.html, Último acesso em Outubro de 2016.

Project, X. (2013). Xenproject. https://www.xenproject.org/, Último acesso em Outubro de 2016.

Sandro Melo, Cesar Domingos, L. C. e. T. M. (2006). *BS7799: Da prática à tática em servidores Linux*. Editora Alta Books Ltda, São Paulo.

Santos, T. (2015). Vantagens de uma hospedagem vps para os seus negócios. https://www.impreza.com.br/blog/vantagens-de-uma-hospedagem-vps-para-os-seus-segocios/, Último acesso em Novembro de 2016.

Squid (2016). Squid. http://www.squid-cache.org/, Último acesso em Novembro de 2016.

Titon, R. (2013). Segurança da informação aplicada a servidores utilizando técnicas de hardening. Master's thesis, UFSM, http://repositorio.ufsm.br:8080/xmlui/handle/1/249.

Venema, W. (2016). Postfix. http://www.postfix.org/, Último acesso em Outubro de 2016.

Wireshark (2016). Wireshark. https://www.wireshark.org/, Último acesso em Outubro de 2016.

Experimentos Numéricos Para o Problema da Coloração Orientada em Grafos Cúbicos

Mateus de Paula Ferreira[1], **Hebert Coelho da Silva**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74690-900 – Goiânia – GO – Brasil

{mateusferreira, hebert}@inf.ufg.br

Abstract. *Let $\overrightarrow{G} = (V, A)$ be an oriented graph, $\overrightarrow{(x,y)}$, $\overrightarrow{(z,t)} \in A(\overrightarrow{G})$, and C a set of k colors, a function $c : C \to \overrightarrow{G}$ such that $c(x) \neq c(y)$ and if $c(x) = c(t)$ then $c(y) \neq c(z)$ is called oriented k-coloring (Introduced by [Raspaud and Sopena 1994]). In [Sopena 1997] Sopena conjecture that if $\overrightarrow{G}$ is an oriented graph with $\Delta(\overrightarrow{G}) \leq 3$ then $\chi_o(\overrightarrow{G}) \leq 7$. In this work we present results of numerical experiments that strengthen the Sopena's conjecture, also we introduce a approximation algorithm for the oriented coloring problem and we compare with the approximation algorithm of Culus and Demange [Culus and Demange 2006].*

Resumo. *Seja $\overrightarrow{G} = (V, A)$ um grafo orientado, $\overrightarrow{(x,y)}$, $\overrightarrow{(z,t)} \in A(\overrightarrow{G})$ e C um conjunto com k cores, uma função $c : C \to \overrightarrow{G}$ tal que $c(x) \neq c(y)$ e se $c(x) = c(t)$ então $c(y) \neq c(z)$ é chamada de k-coloração orientada (Introduzido por [Raspaud and Sopena 1994]). Em [Sopena 1997] Sopena conjecturou que se $\overrightarrow{G}$ é um grafo orientado com $\Delta(\overrightarrow{G}) \leq 3$ então $\chi_o(\overrightarrow{G}) \leq 7$. Neste trabalho apresentamos resultados de experimentos numéricos que fortalecem a conjectura de Sopena. Também introduzimos um algoritmo de aproximação para o problema da coloração orientada e o comparamos com o algoritmo de aproximação de Culus e Demange [Culus and Demange 2006].*

1. Introdução

Neste trabalho serão utilizadas as notações e terminologias padrão em teoria dos grafos. O par $G = (V, E)$ é um *grafo simples*, onde $V(G)$ é um conjunto de vértices e $E(G)$ um conjunto de arestas, não sendo permitido arestas paralelas ou laços. Um *grafo orientado* $\overrightarrow{G} = (V, A)$ é obtido de G definindo uma orientação para cada aresta de $E(G)$, esta orientação é chamada de *arco*, e G é chamado de grafo *subjacente* de $\overrightarrow{G}$. Um *multigrafo misto* é uma tripla $M = (V, A, E)$, que contém arcos e arestas, além disso permite que existam $(x, y) \in E(M)$ e $\overrightarrow{(x,y)} \in A(M)$.

Sejam $\overrightarrow{G}$ e $\overrightarrow{H}$ dois grafos orientados. Um *isomorfismo* entre $\overrightarrow{G}$ e $\overrightarrow{H}$ é definido como uma função bijetora $f : V(\overrightarrow{G}) \to V(\overrightarrow{H})$, em que $\overrightarrow{(x,y)} \in A(\overrightarrow{G})$ se e somente se $\overrightarrow{(f(x), f(y))} \in A(\overrightarrow{H})$. Se $|V(\overrightarrow{G})| = |V(\overrightarrow{H})|$ e existe um isomorfismo entre $\overrightarrow{G}$ e $\overrightarrow{H}$ então $\overrightarrow{G}$ e $\overrightarrow{H}$ são *isomorfos*.

Sejam $\overrightarrow{G}$ e $\overrightarrow{H}$ dois grafos orientados. Um *homomorfismo* entre $\overrightarrow{G}$ e $\overrightarrow{H}$ é uma função $f : V(\overrightarrow{G}) \to V(\overrightarrow{H})$, onde se o arco $\overrightarrow{(x,y)} \in A(\overrightarrow{G})$ então $\overrightarrow{(f(x), f(y))} \in A(\overrightarrow{H})$.

O *grau* de um vértice $x \in V(\overrightarrow{G})$ é definido por $d_{\overrightarrow{G}}(x) = |\{\overrightarrow{(x,y)} \in A(\overrightarrow{G})$ ou $\overrightarrow{(y,x)} \in A(\overrightarrow{G}); \forall y \in V(\overrightarrow{G})\}|$. Os *graus máximo e mínimo* de $\overrightarrow{G}$ são definidos respectivamente por $\Delta(\overrightarrow{G}) = max\{d_{\overrightarrow{G}}(x); x \in V(\overrightarrow{G})\}$ e $\delta(\overrightarrow{G}) = min\{d_{\overrightarrow{G}}(x); x \in V(\overrightarrow{G})\}$.

Seja $x \in V(\overrightarrow{G})$, $\Gamma^+_{\overrightarrow{G}}(x) = \{y; \overrightarrow{(x,y)} \in A(\overrightarrow{G})\}$ é o *conjunto dos sucessores de* x, e $\Gamma^-_{\overrightarrow{G}}(x) = \{y; \overrightarrow{(y,x)} \in A(\overrightarrow{G})\}$ o *conjunto dos predecessores de* x. O conjunto dos sucessores ou predecessores também pode ser definido sobre um conjunto de vértices $U \subset V(\overrightarrow{G})$, onde $\Gamma^+_{\overrightarrow{G}}(U) = \bigcup_{x \in U} \Gamma^+_{\overrightarrow{G}}(x)$ e $\Gamma^-_{\overrightarrow{G}}(U) = \bigcup_{x \in U} \Gamma^-_{\overrightarrow{G}}(x)$. O subgrafo induzido pelo conjunto U é definido por $\overrightarrow{G}[U]$, onde $V(\overrightarrow{G}[U]) = U$ e $A(\overrightarrow{G}[U]) = \{\overrightarrow{(x,y)} \in A(\overrightarrow{G}) \Leftrightarrow x, y \in U\}$.

Conjuntos de grafos que compartilham certas propriedades recebem nomes especiais, por exemplo, um grafo G em que todo vértice tem grau três é dito *cúbico*. Um grafo orientado $\overrightarrow{G}$ é cúbico se o seu grafo subjacente é cúbico. O grafo *completo* com n vértices denotado por K_n é aquele em que cada vértice é adjacente a todos os outros vértices. A orientação de um grafo completo é um *torneio*, sendo denotado por $\overrightarrow{T_n}$.

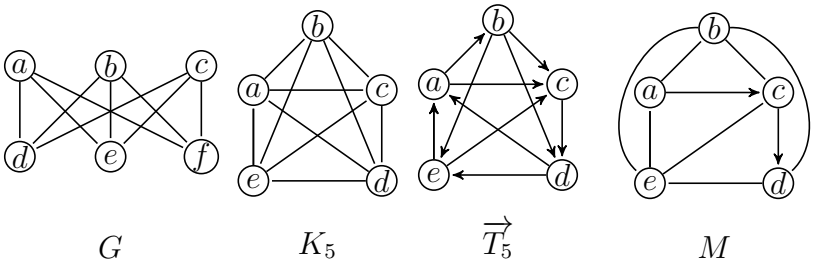


Figura 1. G um grafo cúbico com 6 vértices; O grafo completo K_5; Um torneio $\overrightarrow{T_5}$, uma das possíveis orientações para K_5; Um multigrafo misto M.

Na Figura 1 temos $d_{\overrightarrow{T_5}}(a) = 4$, $\Delta(\overrightarrow{T_5}) = 4$ e $\delta(\overrightarrow{T_5}) = 4$. Seja $U = \{a, d\} \in V(\overrightarrow{T_5})$ $\Gamma^+_{\overrightarrow{T_5}} = \{a, b, c, e\}$, $\Gamma^-_{\overrightarrow{T_5}} = \{b, c, d, e\}$.

O problema da *k-coloração orientada* foi introduzido na literatura por [Raspaud and Sopena 1994]. A k-coloração orientada de um grafo orientado $\overrightarrow{G}$ é a partição de $V(\overrightarrow{G})$ em k conjuntos de cores, tal que dois vértices adjacentes nunca pertençam ao mesmo subconjunto. Além disso, todos os arcos que conectam quaisquer dois subconjuntos de cores devem ter a mesma direção. Em outras palavras a coloração orientada segue as regras:

1. Se $\overrightarrow{(x,y)} \in A(\overrightarrow{G})$, então $cor(x) \neq cor(y)$.
2. Se $\overrightarrow{(x,y)}, \overrightarrow{(z,t)} \in A(\overrightarrow{G})$ e $cor(x) = cor(t)$ então $cor(y) \neq cor(z)$.

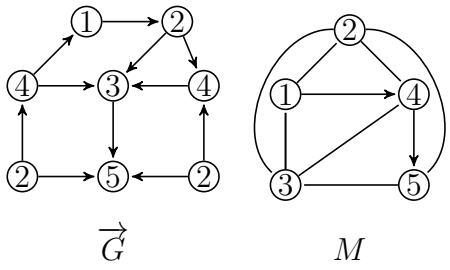
O problema da k-coloração orientada de um grafo orientado $\overrightarrow{G}$ pode ser visto como um homomorfismo de $\overrightarrow{G}$ em um grafo orientado $\overrightarrow{H}$ com k vértices, onde os vértices de $\overrightarrow{H}$ representam as cores utilizadas para colorir $\overrightarrow{G}$. Essa relação do problema da coloração com um homomorfismo já foi estudada para o caso não orientado em [Hell and Nešetřil 1990].

O *número cromático orientado*, denotado por $\chi_o(\overrightarrow{G})$, é o menor *k* tal que $\overrightarrow{G}$ admite uma *k*-coloração orientada.

A k*-coloração mista* de um multigrafo misto M, apresentado por [Culus and Demange 2006], é a partição de $V(M)$ em k classes de cores com valores numéricos {1, 2, *i*, . . ., k} onde $1 \leq i \leq k$, definida pela função *l*:

- $l : V(M) \rightarrow \{1, 2, i, \ldots, k\}$; $1 \leq i \leq k$
- $cor(x) \neq cor(y)$ se *(x, y)* $\in E(M)$
- $cor(x) < cor(y)$ se $\overrightarrow{(x, y)} \in A(M)$

Na Figura 2 exibimos exemplos de coloração orientada e coloração orientada mista.



$\overrightarrow{G}$ M

Figura 2. Uma coloração orientada de um grafo orientado $\overrightarrow{G}$ de 8 vértices, $\chi_o(\overrightarrow{G}) = 5$; 5-coloração orientada mista para M.

A complexidade do problema do número cromático orientado (OCN_k) já foi bastante estudada na literatura, sendo definida por:

$$\textbf{OCN}_k \begin{cases} \text{Número orientado cromático} \leq k\text{, onde } k \text{ é um inteiro positivo} \\ \text{Entrada: Um grafo orientado } \overrightarrow{G} \text{ e um inteiro positivo } k. \\ \text{Problema: Existe uma } k\text{-coloração orientada para } \overrightarrow{G}? \end{cases}$$

De acordo com [Klostermeyer and MacGillivray 2004] o problema OCN_k é polinomial para $k \leq 3$ e NP-completo para $k \geq 4$. Assim, de maneira geral, o problema é NP-completo. A literatura do número cromático orientado é extensa, um resumo sobre o assunto pode ser visto em [Sopena 2015].

Em [Coelho et al. 2015] foi demonstrado que se $\overrightarrow{G}$ é conexo, planar e cúbico, OCN_k ainda é NP-completo para $k > 3$. Complementando o resultado, em [Coelho et al. 2016], foi apresentado que se $\overrightarrow{G}$ é conexo, planar, bipartido, acíclico e com o grau máximo igual a três, OCN_k é NP-completo para $k > 3$. Sopena em [Sopena 1997] mostrou que $\chi_o(\overrightarrow{G}) \leq 16$ onde o grafo subjacente G de $\overrightarrow{G}$ é conexo e $\Delta(\overrightarrow{G}) \leq 3$, e também conjecturou que $\chi_o(\overrightarrow{G}) \leq 7$. Neste mesmo trabalho Sopena também demonstrou que se o grau máximo de um grafo orientado $\overrightarrow{G}$ é menor do que três é possível colorir $\overrightarrow{G}$ em tempo polinomial. Argumentos que fortalecem a conjectura de Sopena podem ser encontrados em [da Silva 2013].

Um algoritmo de aproximação polinomial para um problema de otimização é definido como um algoritmo de tempo polinomial para qualquer instância do problema, que produz uma solução com um valor entre um fator de aproximação α e a solução ótima [Williamson and Shmoys 2011].

A conjectura de Sopena em relação ao limite inferior do número cromático orientado de grafos cúbicos é um problema em aberto desde 1997. O problema da coloração orientada OCN_k tem diversas aplicações, sendo possível modelar problemas reais utilizando OCN_k. A modelagem de um problema utilizando OCN_k é apresentada em [Culus and Demange 2006].

Neste trabalho apresentamos resultados de experimentos numéricos que fortalecem a conjectura de Sopena. Tais resultados foram obtidos da implementação de quatro algoritmos para resolver OCN_k. Na Seção 2 apresentamos dois algoritmos de força bruta, o primeiro algoritmo elaborado e implementado por nós, gera o número cromático orientado de qualquer grafo orientado $\overrightarrow{G}$ e utilizamos seus resultados como base de comparação para os algoritmos de aproximação da Seção 3. O segundo algoritmo da Seção 2 (obtido de [da Silva 2013]) verifica se é possível colorir um grafo orientado $\overrightarrow{G}$ com até 7 cores.

Na seção 3 desenvolvemos e implementamos o Algoritmo 3 de aproximação que utiliza a abordagem gulosa, e o comparamos com a nossa implementação do algoritmo de Culus e Demange [Culus and Demange 2006]. Por fim, apresentamos nossas conclusões e trabalhos futuros. Os algoritmos foram implementados utilizando a linguagem C++ e os grafos armazenados em matrizes de adjacências. Utilizamos um computador que tem um processador de 1.8GHz e 8 gigas de memória RAM para a execução dos algoritmos. Os arquivos com códigos fonte das implementações dos algoritmos e também os arquivos com os resultados computacionais estão disponíveis em <goo.gl/GEp2eX>.

2. Força Bruta

Como descrito na introdução, o problema da coloração orientada é NP-completo mesmo quando restrito a grafos cúbicos. Assim, para gerar os experimentos numéricos que corroborem com a conjectura de Sopena, utilizamos a estrategia de força bruta para encontrar o menor número de cores em grafos cúbicos. Obtivemos todos os grafos cúbicos de 4 a 14 vértices em [Brinkmann et al. 2013].

Nesta seção serão apresentados dois algoritmos de força bruta, o Algoritmo 1 para achar o número cromático orientado de qualquer grafo orientado $\overrightarrow{G}$ e o Algoritmo 2 para verificar se é possível colorir $\overrightarrow{G}$ com 7 cores. Devido aos nossos experimentos numéricos, ambas implementações serviram para fortalecer a conjectura de Sopena.

2.1. Algoritmo para o Caso Geral

O Algoritmo 1 gera o número cromático orientado de um grafo orientado $\overrightarrow{G}$. Este algoritmo será utilizado como base de comparação em relação ao número de cores e tempo de execução para os algoritmos de aproximação. Nas seções subsequentes faremos uma análise de sua complexidade e serão apresentados resultados computacionais. No Algoritmo 1, $C_{rep}(n, c)$ representa o número de combinações com repetição de tamanho n com c objetos, u é um vetor de tamanho n que sera utilizado para armazenar a i-ésima

combinação, pois é inviável armazenar todas as combinações. O número c representa a quantidade de cores que o algoritmo tenta utilizar a cada iteração do laço da linha 1.

Algoritmo 1 *Força Bruta 1*

Entrada: Um grafo orientado $\overrightarrow{G}$ com n vértices.
Saída: $\chi_o(\overrightarrow{G})$.

```
1:  para c = 1 até n faça
2:      χ_o(G→) ← n + 1;
3:      para i = 1 até C_rep(n, c) faça
4:          u ← i-ésima combinação com repetição de tamanho n com c objetos;
5:          Atribua a cor u_i ∈ u para v_i ∈ V(G→), 1 ≤ i ≤ n;
6:          se G→ tem uma coloração orientada válida então
7:              χ_o(G→) ← c;
8:          fim se
9:      fim para
10:     se χ_o(G→) ≠ n + 1 então
11:         retorne χ_o;
12:     fim se
13: fim para
```

2.1.1. Análise de Complexidade de Tempo

Nossa análise terá o foco nos trechos mais significativos do algoritmo. Seja $n = |V(\overrightarrow{G})|$, $m = |A(\overrightarrow{G})|$, e c um número de cores. Primeiro calculamos o número de combinações com repetição $C_{rep}(n, c)$ de c cores em n vértices:

$$C_{rep}(n, c) = C(n + c - 1, c)$$
$$C_{rep}(n, c) = \frac{(n+c-1)!}{c!(n-1)!}$$

A linha 3 é um laço que se repete o número de combinações com repetição de tamanho n para c objetos, assim tem custo $O(C_{rep}(n, c))$.

Na linha 6 é verificado se $\overrightarrow{G}$ tem una coloração válida. Considerando uma estratégia simples de verificação temos que comparar cada aresta com todas as outras arestas, o que tem uma complexidade de $O(m \cdot (m-1))$. Assim a complexidade total no pior caso do Algoritmo 1 é $O(n \cdot C_{rep}(n, c) \cdot m \cdot (m-1))$, ou seja, é fatorial.

2.1.2. Resultados Computacionais

O conjunto de dados que iremos utilizar para a execução do algoritmo de força bruta foi composto por todos grafos cúbicos de 4 a 14 vértices. O tempo médio em segundos apresentado na Tabela 1 representa a execução de todas as 2^m orientações de todos os grafos cúbicos para aquele número de vértices. O tempo médio gasto a partir de 12 vértices é resultado de uma aproximação, pois a quantidade de tempo que levaríamos para executar tais entradas é inviável para esta estrategia de força bruta.

Tabela 1. Execuções do Algoritmo *Força Bruta 1*.

N^o de vértices	N^o de Cúbicos	N^o de Orientações	Tempo Médio
4	1	64	0.000
6	2	1024	1.090
8	5	20480	63.990
10	19	622592	7955.984
12	85	$2.23 \cdot 10^7$	66900000.481
14	509	$1.07 \cdot 10^9$	79180000000.503

O número cromático orientado de nenhum dos grafos testados pelo Algoritmo 1 foi maior do que 7, sendo assim um argumento favorável a conjectura de Sopena.

2.2. Torneios Proibidos

Nesta seção temos o objetivo de continuar testando a conjectura de Sopena de que $\chi_o(\overrightarrow{G}) \leq 7$ para todo $\overrightarrow{G}$ que é conexo e $\Delta(\overrightarrow{G}) = 3$. Os resultados apresentados aqui tem como base o trabalho de [da Silva 2013], que usa o algoritmo OCN_7 para verificar que $\chi_o(\overrightarrow{G}) \leq 7$ para todo grafo orientado $\overrightarrow{G}$ que é cúbico e $|V(\overrightarrow{G}|) \leq 18$, sendo este um resultado favorável a conjectura.

Segundo [Gross and Yellen 2004], existem 456 torneios de 7 vértices não isomorfos. Tais torneios podem ser encontrados na base de dados de [McKay 2013]. Nós os indexamos de 0 a 455.

Algoritmo 2 OCN_7

Entrada: Um grafo G com n vértices e m arestas.

Saída: Uma 7 coloração orientada para toda orientação $\overrightarrow{G}^i$ de G, ou um grafo orientado $\overrightarrow{G}^i$ com 8 cores.

1: **para** $i = 0$ **até** $2^m - 1$ **faça**
2: $flag \leftarrow 0$;
3: Gerar orientação $\overrightarrow{G^i}$ de G;
4: **para** $j = 0$ **até** 455 **faça**
5: **se** $existe\ homomorfismo\ de\ \overrightarrow{G}^i\ para\ \overrightarrow{T}_7^j$ **então**
6: Armazene $\overrightarrow{T_7^j}$;
7: $flag \leftarrow j$;
8: $j \leftarrow 455 + 1$
9: **fim se**
10: **fim para**
11: **se** $flag = 0$ **então**
12: **retorne** $\overrightarrow{G^i}$;
13: **fim se**
14: **fim para**

O algoritmo gera todas as 2^m orientações possíveis para o grafo cubico de entrada, e verifica se existe um homomorfismo de cada orientação para algum dos 456 torneios de 7 vértices. Se existe pelo menos um grafo cúbico orientado que não tem homomorfismo

com o torneio $\overrightarrow{T_7^j}$, então $\overrightarrow{T_7^j}$ é um torneio que não serve para colorir toda a classe dos grafos cúbicos orientados e denominamos $\overrightarrow{T_7^j}$ de *torneio proibido*.

O objetivo de encontrar torneios proibidos é de diminuir o número de torneios que ainda tem a possibilidade de ser a palheta de cores da classe e tornar possível a busca de um padrão estrutural nestes torneios remanescentes que possibilite a prova ou a refutação da conjectura de Sopena.

2.2.1. Análise de Complexidade de Tempo

O algoritmo OCN_7 também utiliza a abordagem da força bruta, mas como seu proposito é diferente e temos parâmetros adicionais definidos (como o número de cores fixado em 7), sua complexidade é menor do que a do algoritmo do caso geral. O que podemos observar comparando os tempos de execução das Tabelas 2 e 1.

O laço da linha 1 enumera todas orientações com isomorfismo para o grafo orientado $\overrightarrow{G}$, com n vértices e m arestas, assim seu tempo é 2^m.

Gerar uma orientação $\overrightarrow{G^i}$ de G tem uma complexidade de $O(n^2)$. Para verificar se é possível gerar um homomorfismo de $\overrightarrow{G^i}$ para $\overrightarrow{T_7^j}$ temos que gerar $C_{rep}(n, 7)$, mas a diferença é que não precisamos encontrar todas as combinações que são válidas, apenas a primeira, mas no pior caso quando não existe uma função de homomorfismo todas as combinações ainda serão verificadas. Com isso, sabendo que verificar se existe um homomorfismo de $\overrightarrow{G^i}$ para $\overrightarrow{T_7^j}$ é $O(m \cdot (m-1))$, no pior caso o laço da linha 4 toma $O(456 \cdot C_{rep}(n, 7) \cdot m \cdot (m-1))$, e o algoritmo total tem $O((2^m) \cdot 456 \cdot C_{rep}(n, 7) \cdot m \cdot (m-1))$.

2.2.2. Resultados Computacionais

A Tabela 2 trás os resultados das execuções de OCN_7, para os grafos cúbicos de 4 a 14 vértices. O tempo médio foi calculado com a diferença de que na linha 4 em vez de executarmos todos os 456 torneios escolhemos apenas o *torneio mágico* $\overrightarrow{T}_7^{114}$, veja Figura 3, pois [da Silva 2013] mostra que é garantido que este torneio colore todos os grafos cúbicos com até 18 vértices. O tempo médio em segundos da Tabela 2 representa o tempo total de execução de todas as orientações de todos os grafos cúbicos para aquele número de vértices.

Tabela 2. Execução de OCN_7 fixando o torneio $\overrightarrow{T}_7^{114}$ para todas as orientações.

N^o de vértices	N^o de Cúbicos	N^o de Orientações	Tempo Médio
4	1	64	0.000
6	2	1024	0.000
8	5	20480	1.012
10	19	622592	39.701
12	85	$2.23 \cdot 10^7$	1745.007
14	509	$1.07 \cdot 10^9$	109800.900

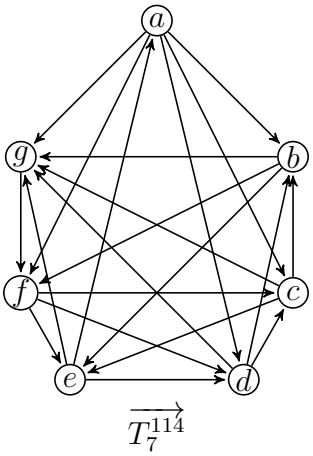


Figura 3. O torneio mágico $\overrightarrow{T}_7^{114}$.

Utilizando os resultados de cada execução de OCN_7 para os grafos cúbicos de 4 a 14 vértices, ampliamos os 275 torneios proibidos apresentados em [da Silva 2013], para 431 torneios proibidos, restando apenas 25 (456-431=25) torneios que ainda não foram determinados ou não são proibidos e são apresentados na Tabela 3. Estes resultados corroboram a conjectura de Sopena pois ainda existem torneios de 7 vértices que podem ser a palheta de cores para a classe dos grafos cúbicos. Caso cada um dos 25 torneios apresentados na Tabela 3 não possam colorir algum grafo cúbico, a conjectura de Sopena será invalidada.

Tabela 3. Torneios que não foram determinados ou não são proibidos.

Torneios que não foram determinados ou não são proibidos.
95, 114, 156, 161, 162, 163, 166, 170, 199, 226, 241, 243, 246, 308, 350, 351, 353, 357, 365, 377, 378, 387, 388, 395, 396.

Podemos utilizar OCN_7 como uma abordagem de aproximação para os grafos cúbicos até 18 vértices utilizando o torneio mágico. Apesar de não ter tempo polinomial o fato de não ser necessário encontrar todas as combinações válidas, como foi apresentado na análise de complexidade, faz com que seus tempos de execução sejam consideravelmente menores.

Apesar do Algoritmo 2 OCN_7 ser de força bruta, ele pode não gerar uma coloração ótima em alguns casos. O fator de aproximação α de OCN_7 para todos os grafos cúbicos de 4 a 10 vértices representa a distância da solução aproximada da solução ótima. A maneira de calcular α é apresentada a seguir.

Considere que $m_A(\overrightarrow{G})$ representa o número de cores utilizadas por OCN_7 para colorir $\overrightarrow{G}$, m o número de arestas de $\overrightarrow{G}$, $\overrightarrow{G_i}$ a i-ésima orientação de $\overrightarrow{G}$, e k o número de grafos cúbicos não orientados para a determinada quantidade de vértices.

$$\alpha^j = \frac{\sum_{i=0}^{2^m-1} \frac{m_A(\overrightarrow{G_i})}{\chi_o(\overrightarrow{G_i})}}{2^m-1}$$

$$\alpha = \frac{\sum_{j=0}^{k} \alpha^j}{k}$$

Tabela 4. Fator de aproximação α de OCN_7 para os grafos cúbicos de 4 a 10 vértices.

N^o de vértices	α
4	1.000000
6	1.056006
8	1.152894
10	1.207796

O fator de aproximação α considerado é uma média simples da diferença entre o número de cores obtido pelo algoritmo OCN_7 e o número cromático obtido pelo Algoritmo 1 para todos os grafos cúbicos com n vértices, $4 \leq n \leq 10$.

3. Algoritmos de Aproximação

A abordagem de força bruta é inviável quando aumentamos o número de vértices de $\overrightarrow{G}$. Algoritmos de aproximação são uma alternativa com a finalidade de gerar soluções aproximadas com tempo polinomial para OCN_k, quando o número de vértices de $\overrightarrow{G}$ cresce.

Nesta seção apresentamos o Algoritmo 3 que desenvolvemos e implementamos, e comparamos os resultados com a nossa implementação de GOC (até o momento de nossa revisão bibliográfica a primeira implementação conhecida) que foi apresentado por [Culus and Demange 2006].

3.1. Abordagem Gulosa

Nosso algoritmo passa por cada vértice uma única vez e o colore com a menor cor disponível no momento. A disponibilidade de uma cor c para um vértice v é verificada checando as adjacências de v e se existe um caminho de tamanho 2 para algum vértice que tenha a mesma cor que v ou um caminho de tamanho 2 que parta de algum vértice $y \in V(\overrightarrow{G})$ que termine em v e que $cor(v) = cor(y)$. Também checa se é possível formar esses caminhos indiretamente utilizando outros vértices que tenham a mesma cor.

3.1.1. Análise de Complexidade de Tempo

Considere $n = |V(\overrightarrow{G})|$ e $m = |A(\overrightarrow{G})|$. Na linha 5 temos a verificação se a $corAtual$ é igual a cor de alguma das adjacências, o que tem a complexidade de $O(\Gamma^+_G(v) \cup \Gamma^-_G(v))$ o que no pior caso é $O(n-1)$. Na linha 8 e na linha 11 temos a verificação se existe um caminho de tamanho 2 chegando ou saindo de v diretamente ou indiretamente, o que é $O(2 \cdot (m \cdot (m-1)))$.

O laço da linha 3 será executado n vezes para cada vértice no pior caso, o que faz com que a complexidade total do algoritmo seja $O(n^2 \cdot max\{n-1, 2 \cdot (m \cdot (m-1))\})$, ou seja, o Algoritmo 3 tem tempo polinomial.

Algoritmo 3 *Aproximação Gulosa*

Entrada: Um grafo orientado $\overrightarrow{G}$ com n vértices.
Saída: Uma coloração orientada para $\overrightarrow{G}$.

1: **para todo** $v \in V(\overrightarrow{G})$ **faça**
2: $corAtual \leftarrow 0$;
3: **repita**
4: $flag \leftarrow 0$;
5: **se** $corAtual \in cor(\Gamma^+_G(v) \cup \Gamma^-_G(v))$ **então**
6: $flag \leftarrow 1$;
7: **fim se**
8: **se** $\exists\ z \in V(\overrightarrow{G})$ tal que $[\overrightarrow{(v,z)}, \overrightarrow{(z,y)} \in A(\overrightarrow{G})$ **ou** $\overrightarrow{(y,z)}, \overrightarrow{(z,v)} \in A(\overrightarrow{G})]$ **e** $[corAtual = cor(y)]$ **então**
9: $flag \leftarrow 1$;
10: **fim se**
11: **se** $\exists\ z,t \in V(\overrightarrow{G})$ tal que $[\overrightarrow{(v,y)}, \overrightarrow{(z,t)} \in A(\overrightarrow{G})]$ **e** $[cor(y) = cor(t)]$ **e** $[corAtual = cor(t)]$ **então**
12: $flag \leftarrow 1$;
13: **fim se**
14: **se** $flag = 0$ **então**
15: $cor(v) \leftarrow cor$;
16: **fim se**
17: $corAtual \leftarrow corAtual + 1$;
18: **até** $flag \neq 1$
19: **fim para**

3.2. Abordagem Gulosa de Culus e Demange

Os algoritmos GMC e GOC são apresentados por [Culus and Demange 2006], e consistem em uma generalização para o caso orientado do algoritmo guloso usual para a coloração não orientada. Culus e Demange mostram que uma k-coloração orientada de $\overrightarrow{G}$ é equivalente a uma k-coloração mista de $M(\overrightarrow{G})$ e sua recíproca, e fazem as seguintes definições necessárias para o algoritmo:

- Dado um grafo orientado $\overrightarrow{G}$, definimos o multigrafo misto $M(\overrightarrow{G})$ associado a $\overrightarrow{G}$ por: $V(M(\overrightarrow{G})) = V(\overrightarrow{G})$, $A(M(\overrightarrow{G})) = A(\overrightarrow{G})$ e $E(M(\overrightarrow{G})) = \{\{x,y\}; [\exists z \in V(\overrightarrow{G}); \overrightarrow{(x,z)}, \overrightarrow{(z,y)} \in A(\overrightarrow{G})]$ e $[\exists z \in V(\overrightarrow{G}); \overrightarrow{(y,z)}, \overrightarrow{(z,x)} \in A(\overrightarrow{G})]\}$.
- Dado um multigrafo misto M e um vértice $v \in V(M)$, definimos o conjunto de vértices $B(v,2) = \{y; [(v,y) \in E(M)]$ e $[\overrightarrow{(v,y)} \in A(M)]$ e $[\exists z \in V(M); \overrightarrow{(v,z)}, \overrightarrow{(z,t)} \in A(M)]$ e $[\exists z \in V(M); \overrightarrow{(t,z)}, \overrightarrow{(z,v)} \in A(M)]\}$.

3.2.1. Análise de Complexidade de Tempo

Considere $n = |V(M)|$, $m = |A(M)|$ e $p = |E(M)|$. Começaremos nossa análise pelo algoritmo GMC. A operação $B(v,2)$ é $O(m \cdot (m+p-1))$, para executarmos a instrução da linha 4 no pior caso temos que gerar $B(v,2)$ n vezes, tomando $O(n \cdot (m \cdot (m+p-$

Algoritmo 4 GMC

Entrada: Um multigrafo misto $M = (V, E, A)$.
Saída: Um conjunto independente S de M.

1: $S \leftarrow \emptyset$;
2: $U \leftarrow V$;
3: **enquanto** $U \neq \emptyset$ **faça**
4: Escolha v tal que $|B(v, 2)|$ seja minimo em $M[U]$;
5: $S \leftarrow S \cup \{v\}$;
6: $U \leftarrow U \setminus B(v, 2)$;
7: **fim enquanto**
8: **retorne** S

Algoritmo 5 GOC

Entrada: Um grafo orientado $\overrightarrow{G}$.
Saída: Uma *k*-coloração mista de $M(\overrightarrow{G})$.

1: Construa $M(\overrightarrow{G})$;
2: $U \leftarrow V$;
3: $i \leftarrow 1$;
4: **enquanto** $|U| > 0$ **faça**
5: Escolha no máximo $\log(|U|)$ vértices em $GMC(\overrightarrow{G}[U])$ para a cor i;
6: Seja V_{min} o subconjunto de menor ordem entre $\Gamma^{+}_{G}(GMC(\overrightarrow{G}[U]))$ e $\Gamma^{-}_{G}(GMC(\overrightarrow{G}[U]))$;
7: Todo vértice de V_{min} recebe uma cor diferente em $\{i + 1, \ldots, i + |V_{min}|\}$;
8: $U \leftarrow U \setminus (GMC(\overrightarrow{G}[U]) \cup V_{min})$;
9: $i \leftarrow i + |V_{min}| + 1$;
10: **fim enquanto**

$1)))$, o laço da linha 3 de GMC é executado no máximo n vezes o que faz com que a complexidade de GMC seja $O(n^2 \cdot (m \cdot (m + p - 1)))$.

Sabemos que GMC é chamado $\lfloor \log_3(n) \rfloor$ em GOC [Culus and Demange 2006]. Considerando as chamadas a GMC como a operação de maior custo em GOC a complexidade de GOC é $O(\lfloor \log_3(n) \rfloor \cdot n^2 \cdot (m \cdot (m + p - 1)))$.

3.3. Resultados Computacionais

Nesta seção apresentamos os resultados das implementações dos Algoritmos 3 e 5. A Tabela 5 contém o tempo médio das execuções dos Algoritmos 3 e 5 para os grafos cúbicos de 4 a 14 vértices. O Tempo Médio representa o tempo de execução de todas as orientações possíveis para cada um dos grafos cúbicos com o determinado número de vértices.

A comparação da Tabela 6 com a Tabela 4 mostra que o fator de aproximação do Algoritmo 3 tem um resultado melhor do que o fator de aproximação de OCN_7, representando que suas soluções estão mais próximas da solução ótima.

A comparação das colunas 4 e 5 da Tabela 5 mostra que os tempos de execução de GOC são menores do que os do Algoritmo 3, mas na Tabela 6 observamos que o fator de

aproximação do Algoritmo 3 tem um crescimento menor do que o de GOC. Os resultados em relação ao tempo são condizentes com a complexidade de cada algoritmo, tendo em vista que GOC tem uma complexidade menor do que a do Algoritmo 3.

Tabela 5. Execuções dos Algoritmos 3 e 5 para os grafos cúbicos de 4 a 14 vértices. NC representa o número de grafos cúbicos com n vértices.

n	NC	N^o de orientações	Tempo Médio Alg. 3	Tempo Médio Alg. 4
4	1	64	0.000	0.000
6	2	1024	0.000	0.000
8	5	20480	1.001	0.894
10	19	622592	12.545	8.301
12	85	$2.23 \cdot 10^7$	707.241	502.559
14	509	$1.07 \cdot 10^9$	49831.018	32347.940

Tabela 6. Fatores de aproximação dos Algoritmos 3 e 5 para os grafos cúbicos de 4 a 10 vértices.

N^o de vértices	α para o Algoritmo 3	α para o Algoritmo 5
4	1.000000	1.000000
6	1.003581	1.185007
8	1.077716	1.310927
10	1.134207	1.432675

4. Conclusão

Neste artigo, apresentamos inicialmente o Algoritmo 1 que utiliza a abordagem da força bruta para encontrar o número cromático orientado de um grafo orientado $\overrightarrow{G}$, este foi utilizado para verificarmos a qualidade das soluções produzidas pelos algoritmos de aproximação. Além disso o Algoritmo 1 serviu para verificar que todos os grafos cúbicos de 4 a 10 vértices tem seu número cromático orientado menor ou igual a 7, corroborando com a conjectura de Sopena. Essa abordagem se mostrou inviável até mesmo para casos em que o número de vértices do grafo é pequeno, tendo em vista que sua execução não foi concluída nem para todos os grafos cúbicos com 12 vértices.

Com o segundo algoritmo apresentado, obtido em [da Silva 2013], nosso objetivo de encontrar outros torneios proibidos para a classe dos grafos cúbicos foi alcançada, visto que conseguimos aumentar o número de torneios proibidos que já eram conhecidos e fornecer mais dados para a análise da conjectura de Sopena. Também como OCN_7 é um algoritmo de força bruta, não foi possível estender nossos testes para um número maior de grafos.

A partir de nossas implementações, percebemos que a utilização de programação paralela (por exemplo, com o uso de placas de vídeo dedicadas e CUDA [Nvidia 2007]) poderemos estender nossos testes, pois utilizando apenas 4 núcleos de uma mesma CPU ao invés de 1 foi possível diminuir os tempos apresentados neste trabalho pela metade. Com isso sugerimos para trabalhos futuros a implementação de OCN_7 utilizando programação paralela, o que pode resultar no aumento do número de torneios proibidos

conhecidos ou até mesmo mostrar que todos os torneios de 7 vértices são proibidos o que refutaria a conjectura de Sopena.

Apresentamos dois algoritmos de tempo polinomial que geram soluções aproximadas para o problema da k-coloração orientada, o Algoritmo 3 que criamos e implementamos tem um melhor fator de aproximação (gera soluções mais próximas da solução ótima) do que o Algoritmo 5 obtido de [Culus and Demange 2006] conforme a Tabela 6, porém o Algoritmo 5 tem uma complexidade um pouco menor. Outro ponto relevante é que o Algoritmo 3 tem um fator de aproximação melhor que o Algoritmo 5 GOC que é polinomial, veja Tabelas 4 e 6.

A demonstração por indução dos limites superiores do número cromático orientado para um grafo $\overrightarrow{G}$ com $\Delta(\overrightarrow{G}) \leq 3$ e conexo, $\chi_o(\overrightarrow{G}) \leq 16$ [Sopena 1997] e $\chi_o(\overrightarrow{G}) \leq 11$ [Sopena and Vignal 1996] talvez possam ser utilizados para criar algoritmos com uma abordagem de aproximação para grafos cúbicos. Sugerimos estes como trabalhos futuros. A formulação de outro algoritmo de aproximação específico para os grafos cúbicos e que consiga gerar resultados favoráveis em relação ao tempo e ao número de cores utilizadas com relação aos Algoritmos 3 e 5 é outra sugestão para trabalhos futuros. O resultado $\chi_o(\overrightarrow{G}) \leq 11$ [Sopena and Vignal 1996] não foi publicado mas supomos que esteja correto e por isso optamos por citá-lo apenas nesta conclusão.

Referências

[Brinkmann et al. 2013] Brinkmann, G., Coolsaet, K., Goedgebeur, J., and Mélot, H. (2013). House of graphs: A database of interesting graphs. *Discrete Applied Mathematics*, 161(1):311–314.

[Coelho et al. 2015] Coelho, H., Faria, L., Gravier, S., and Klein, S. (2015). Complexity of the oriented coloring in planar, cubic oriented graphs. *Matemática Contemporânea*, 44:1–8.

[Coelho et al. 2016] Coelho, H., Faria, L., Gravier, S., and Klein, S. (2016). Oriented coloring in planar, bipartite, bounded degree 3 acyclic oriented graphs. *Discrete Applied Mathematics*, 198:109–117.

[Culus and Demange 2006] Culus, J.-F. and Demange, M. (2006). Oriented coloring: Complexity and approximation. In *SOFSEM 2006: Theory and Practice of Computer Science*, pages 226–236. Springer.

[da Silva 2013] da Silva, H. C. (2013). *Coloração Orientada: Uma Abordagem Estrutural e de Complexidade*. PhD thesis, Universidade Federal do Rio de Janeiro.

[Gross and Yellen 2004] Gross, J. L. and Yellen, J. (2004). *Handbook of graph theory*. CRC press.

[Hell and Nešetřil 1990] Hell, P. and Nešetřil, J. (1990). On the complexity of h-coloring. *Journal of Combinatorial Theory, Series B*, 48(1):92–110.

[Klostermeyer and MacGillivray 2004] Klostermeyer, W. F. and MacGillivray, G. (2004). Homomorphisms and oriented colorings of equivalence classes of oriented graphs. *Discrete Mathematics*, 274(1):161–172.

[McKay 2013] McKay, B. (2013). Combinatorial data. *URL: http://cs. anu. edu. au/bdm/-data/graphs. html*.

[Nvidia 2007] Nvidia, C. (2007). Compute unified device architecture programming guide.

[Raspaud and Sopena 1994] Raspaud, A. and Sopena, E. (1994). Good and semi-strong colorings of oriented planar graphs. *Information Processing Letters*, 51(4):171–174.

[Sopena 1997] Sopena, E. (1997). The chromatic number of oriented graphs. *Journal of Graph Theory*, 25(3):191–205.

[Sopena 2015] Sopena, E. (2015). Homomorphisms and colourings of oriented graphs: An updated survey. *Discrete Mathematics*.

[Sopena and Vignal 1996] Sopena, E. and Vignal, L. (1996). A note on the oriented chromatic number of graphs with maximum degree three. *Rapport de recherche, LaBRI*.

[Williamson and Shmoys 2011] Williamson, D. P. and Shmoys, D. B. (2011). *The design of approximation algorithms*. Cambridge university press.

Um *Gateway* Dinâmico XMPP para Rede de Sensores sem Fio

Thaís Mombach, Bruno Silvestre[1]

[1]Instituto de Informática
Universidade Federal de Goiás – Goiânia, GO – Brazil

thais@inf.ufg.br, brunoos@inf.ufg.br

Resumo. *Este trabalho apresenta uma arquitetura para um* gateway *que permite acesso via RPC a nós de Redes de Sensores Sem Fio (RSSF). O* gateway *fornece meios para descoberta de serviços e a adição dinâmica dos procedimentos remotos. Esses mecanismos facilitam a gerência, pois não há necessidade de modificação no código do* gateway *para acomodar alterações nos serviços oferecidos pela RSSF. Aplicações também podem, dinamicamente, perceber as mudanças e se adaptarem.*

Abstract. *This paper presents an architecture for a gateway that allows access via RPC to nodes of a Wireless Sensor Network (WSN). The gateway provides means for services discovery and the addition of remote procedures dynamically. These mechanisms facilitate the management because there is no need to change the gateway's code to accommodate the changes in WSN. Applications also can dynamically see the changes and adapt itself.*

1. Introdução

Uma rede de sensores sem fio (RSSF) é composta por pequenos equipamentos (nós) com poder computacional e memória limitados. Esses equipamentos são muitas vezes dotados de sensores, o que permite a utilização desse tipo de rede no monitoramento ambiental, por exemplo, ou mesmo de seres humanos [Yick et al. 2008].

A comunicação é feita através de rádio e, devido ao pouco alcance que os rádios possuem, a informação geralmente tem que roteada pelos próprios nós para que ela flua pela rede. Um nó especial chamado de sorvedouro (ou *sink*) tem o papel de conectar a RSSF com outras redes. É comum o nó sorvedouro estar conectado a um computador por um cabo serial e servir como ponto intercomunicação com a Internet.

O conceito de Internet das Coisas (IoT — *Internet of Things*) surgiu como uma evolução do conceito de RSSF, mas com a ideia de que os equipamentos estariam conectados diretamente a Internet, podendo ser acessados sem intermediários [Atzori et al. 2010]. No entanto, a conexão com a Internet tem seu preço. O nó tem que ter um requisito mínimo, principalmente de armazenamento, pois a implementação da pilha IP demanda uma quantidade razoável de recurso. Em [Silvestre et al. 2013] mostramos que um programa desenvolvido em TinyOS [Kaliski 2005] utilizando o protocolo 6LoWPAN (IPv6 para RSSF) para conexão direta com a Internet era maior do que o armazenamento disponível em nós MicaZ[1].

[1]MicaZ possui 4Kb de RAM, processador ATmega128L de 8bits a 8MHz e rádio padrão 802.15.4.

Além disso, temos aplicações legadas que muitas vezes não serão migradas para esse novo modelo de conexão com a Internet. E como essas aplicações geralmente possuem o seu próprio protocolo interno de comunicação, o computador ligado ao nó sorvedouro (*gateway*) fica como um ponto de interconexão e conversão do protocolo interno para a Internet. Dessa forma, o *gateway* deve conhecer o formato interno das mensagens da rede RSSF para exportá-las para a Internet.

Surge então outra questão: qual será a forma de acesso às funcionalidades da RSSF que o *gateway* fornecerá?

Uma abordagem que pode ser empregada é o uso do protocolo XMPP (*Extensible Messaging and Presence Protocol*) [Saint-Andre 2011]. Esse protocolo é orientado a mensagens XML (*Extensible Markup Language*) e, segundo a especificação, permite a troca de mensagens quase em tempo real.

No modelo XMPP pode existir um servidor ou uma federação de servidores de autenticação, mas o objetivo é que a comunicação entre as partes seja baseada em *peer-to-peer*. XMPP também possui um conjunto de extensões que são padronizadas pela *XMPP Standards Foundation*[2]. As extensões utilizam o modelo básico de troca de mensagens e adicionam novas funcionalidades, como por exemplo Serviço de Descoberta ou *Publish-Subscribe*.

O objetivo deste trabalho é construir um *gateway* que forneça aos clientes (aplicações) XMPP acesso aos dados dos sensores da RSSF. Uma característica delineada para o *gateway* é que ele possa ser reconfigurado dinamicamente por um administrador para se adequar aos diferentes formatos das mensagens internas da RSSF. Desse modo, a própria RSSF pode ser modificada, tendo o formato das mensagens alterado, e o *gateway* não necessitaria de ser recompilado para acomodar a nova tradução de mensagens.

Além disso, temos como parte da proposta que os clientes possam descobrir dinamicamente quais funcionalidades são fornecidas pela RSSF, ou seja, a descoberta dos serviços da rede de sensores. Isso possibilita a composição de serviços em tempo de execução e também a adaptação dinâmica do cliente às mudanças da rede.

O restante deste texto está organizado da seguinte forma. Na seção 2 apresentamos como o *gateway* dinâmico foi construído. Na seção 3 descrevemos o teste de validação do *gateway*, juntamente com as ferramentas utilizadas ou desenvolvidas para a realização do teste. Na seção 4 destacamos alguns trabalhos relacionados. Finalmente, na seção 5 apresentamos as conclusões e trabalhos futuros.

2. *Gateway* Dinâmico

A Figura 1 mostra a arquitetura do *Gateway* Dinâmico proposto neste trabalho. O modelo de comunicação entre o *gateway* e a RSSF assumido neste trabalho se baseia em chamada remota de procedimento (RPC). Assim, o cliente envia uma requisição para um nó da RSSF, que vai processar e enviar de volta a resposta (sendo a comunicação intermediada pelo *gateway*). Assim, definimos um conjunto de procedimentos que o cliente pode invocar quando quiser interagir com os nós da RSSF. O módulo de Comandos Ad-hoc é responsável por essa interação de chamadas remotas com o cliente.

[2]http://xmpp.org/extensions/

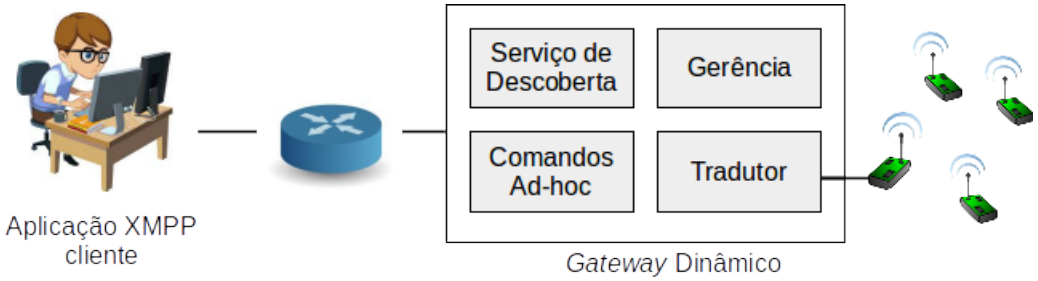


Figura 1. Arquitetura do *Gateway* Dinâmico.

Estamos assumindo que os nós podem prover diferentes funcionalidades, então cada nó pode responder a algum tipo diferente de procedimento. Para auxiliar o cliente no uso, disponibilizamos um módulo de Serviço de Descoberta para que a aplicação cliente possa identificar quais nós fornecem a informação (respondem ao comando) que ela está interessada. O serviço não é tão poderoso como o serviço de *trading* de CORBA [Bolton 2001] ou o UDDI dos *webservices* [Cerami 2002], mas fornece um certo dinamismo à aplicação, até para reagir à mudanças na rede de sensores.

Dependendo do procedimento, ele pode requisitar parâmetros de entrada e pode também retornar ou não valores para o cliente. O *gateway* deve realizar esse mapeamento entre as mensagens XMPP e o formato interno das mensagens trocadas na RSSF. Essa tarefa é realizada pelo módulo Tradutor. Além disso, esse módulo é responsável pela comunicação com o nó sorvedouro da RSSF por meio da porta serial do computador.

O módulo Gerência é utilizado pelo administrador para gerenciar os procedimentos. Internamente, esse módulo deve contactar os demais módulos para que eles atualizem suas informações. Por exemplo, se a RSSF for atualizada e novos procedimentos estão disponíveis em alguns nós, o administrador pode cadastrar esses novos procedimentos, informando nome, parâmetros de entrada, retorno e forma de tradução. Com isso, os novos procedimentos serão listados no Serviço de Descoberta, e os módulos de Comandos Ad-Hoc e Tradutor já saberão tratá-los.

2.1. Implementação do *Gateway*

O *gateway* foi implementado em Java utilizando a biblioteca Smack para XMPP [Ignite Realtime 2016]. No lado da RSSF, estamos assumindo que os nós foram programados utilizando o TinyOS [Levis et al. 2005] (um *framework* para programação de RSSF bem conhecido). Essa definição é importante pois temos que saber qual é o formato de mensagem que a RSSF emprega. Por exemplo, cada mensagem do TinyOS, além dos dados, possui um conjunto de campos de controle, dentre eles, um identificador que define qual é o tipo dessa mensagem. O *gateway* tem que saber como montar as mensagens na forma que a rede sem fio espera (e saber ler essas mensagens de volta).

O módulo Gerência é de certa forma o mais simples, pois ele responde a um conjunto de mensagens XMPP definidas para realizar a inclusão, listagem e remoção dos procedimentos, e vinculação de procedimento ao nó.

A Figura 2 mostra a mensagem enviada ao módulo Gerência para a inclusão de um procedimento. É definido o identificador do procedimento (string), os campos que o cliente deve enviar e os valores que serão retornados. Note que são informados dois

tipos para os campos: `type` é o tipo que será apresentado à aplicação cliente e `wsn-type` é o tipo que a RSSF utiliza. `id-request` e `id-response` também são utilizados pela RSSF, mais especificamente, pelo TinyOS para identificar o tipo de cada das mensagens, ou seja, cada procedimento vai gerar tipos de mensagens únicos para requisição e resposta na RSSF usando TinyOS.

```
<iq from="admin@blackbox/Client"
   to="gateway@blackbox/Server" type="set">
  <query xmlns="br:ufg:inf:dyngateway:command:add">

    <command>getTemperature::1.0</command>

    <id-request>25</id-request>

    <request>
      <field>
        <name>unidade</name>
        <type>integer</type>
        <wsn-type>nx_uint8</wsn-type>
      </field>
    </request>

    <id-response>32</id-response>

    <response>
      <field>
        <name>temperatura</name>
        <type>integer</type>
        <wsn-type>nx_uint16</wsn-type>
      </field>
    </response>
  </query>
</iq>
```

Figura 2. Exemplo de mensagem para adicionar procedimento.

As demais mensagens para o módulo Gerência são simples e não serão mostradas, pois a remoção só indica o identificador do procedimento, e a listagem solicita ao módulo a lista de identificadores dos procedimentos cadastrados. A vinculação é uma mensagem que contém o identificador do procedimento e o identificador do nó — no TinyOS, cada nó é identificado por um número. Isso significa que o mesmo procedimento pode estar disponível para vários nós diferentes.

O módulo Tradutor é responsável por converter as mensagens recebidas do cliente e enviá-las para a RSSF, e vice-versa. Para realizar essa tradução, o módulo utiliza as informações de tipos do campo que foram informados na inclusão do procedimento. Alguns mapeamentos de tipos são diretos, por exemplo, o tipo `float` que segue o padrão IEEE 754 está disponível em várias plataformas de sensores, (e suportada pelo TinyOS) e nos computadores que os clientes utilizam. Por outro lado, grande parte das plataformas de RSSF utilizam números inteiros com ordem de bytes em *big-endian* enquanto os computadores geralmente utilizam *little-endian* (plataforma Intel). Destacamos que a implementação atual do módulo Tradutor atual é fortemente baseada nos tipos utiliza-

dos pelo TinyOS, pois, novamente, assumimos que foi este o *framework* utilizado para programar os nós da RSSF.

O módulo Comandos Ad-Hoc implementa o modelo RPC usando a extensão XEP-0050 (*Ad-Hoc Commands*) [Miller 2015] do protocolo XMPP. Essa extensão trabalha de forma parecida com os formulários HTTP utilizados na web. O cliente inicia o processo enviando a string de identificação do procedimento que deseja executar. Essa identificação é a junção do identificador do nó da RSSF que irá efetivamente executar o procedimento com o identificador do procedimento cadastrado pelo administrador. O módulo então devolve ao cliente um formulário que deve ser preenchido e retornado. Esse formulário contém os nomes dos campos e os seus respectivos tipos.

A Figura 3 mostra um exemplo de formulário que o cliente recebe. O atributo `node` indica o procedimento ao qual o formulário está relacionado (onde `2` é o identificador do nó e `getTemperature` é o identificador do procedimento) e a *tag* `field` é o campo que o cliente deve preencher. (formulários podem ter vários campos a serem preenchidos pelo cliente)

```
<command xmlns="http://jabber.org/protocol/commands"
       node="wsn-mote/2/getTemperature::1.0"
       sessionid="SM82D0Wq5b8S3Ce"
       status="executing">

  <actions><complete/></actions>

  <x xmlns="jabber:x:data" type="form">
    <field var="unidade" type="integer"></field>
  </x>
</command>
```

Figura 3. Exemplo de formulário enviado à aplicação cliente.

O módulo Serviço de Descoberta foi implementado utilizando a extensão XEP-0030 (*Service Discovery*) [Hildebrand et al. 2008]. Essa extensão exporta informações sobre itens cadastrados. Cada item possui um conjunto de características (*features*) e também pode ter subitens. Assim, o cliente pode, a partir de um item conhecido, navegar na árvore, descobrindo novos itens. Esse módulo então permite que aplicações descubram os procedimentos e os nós da RSSF que o implementam.

A Figura 4 demonstra as árvores que o Serviço de Descoberta exporta para os clientes. A árvore (a) tem como raiz o item *wsn-mote* cujos subitens são os identificadores dos nós que implementam algum procedimento. Estes itens por sua vez possuem subitens que são os nomes dos procedimentos. A árvore (b) faz o mapeamento inverso, ela tem como raiz o item *wsn-procedure* e no seu segundo nível estão os procedimentos disponíveis na RSSF, e sob cada um destes estão os identificadores dos nós que os implementam.

Essas duas formas de descoberta permite, por exemplo, que uma aplicação que deseja saber a temperatura de uma área que está sendo monitorada por uma RSSF entre em contato com o *gateway* e solicite ao módulo Serviço de Descoberta todos os nós que respondem ao procedimento `getTemperature`. Depois, ela pode invocar o procedimento

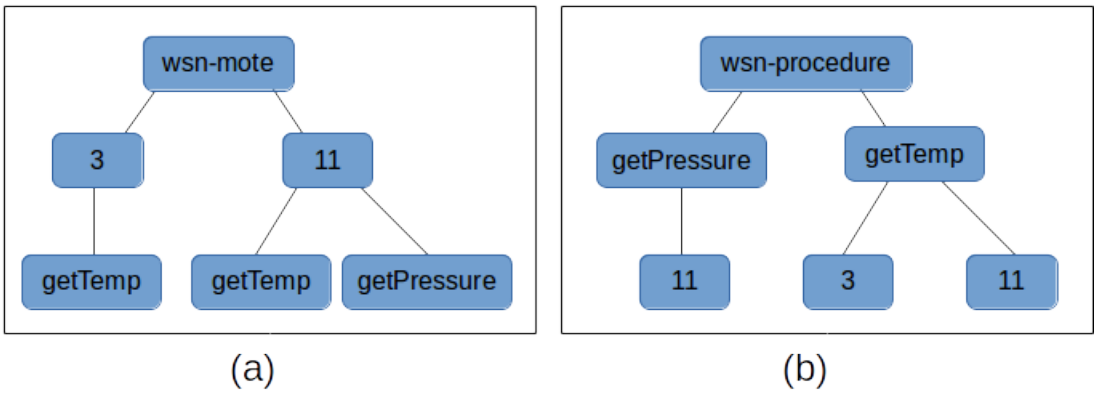


Figura 4. Árvores exportadas pelo módulo Serviço de Descoberta.

em cada um dos nós descobertos. Por outro lado, se a aplicação sabe o posicionamento dos nós, ela pode verificar se os nós que cobrem determinada área implementam o procedimento de leitura da temperatura.

3. Validação da Proposta

Para a validação do *gateway*, utilizamos um conjunto de ferramentas que permitiram demonstrar a dinamicidade proposta.

Utilizamos Terra [Branco et al. 2015], que é uma máquina virtual construída utilizando o TinyOS, para a criação de aplicações para RSSF. Terra permite que o código que é executado nos nós da RSSF sejam atualizados dinamicamente via mensagens sem fio, com isso podemos modificar o funcionamento da RSSF sem a necessidade de desligarmos os nós.

Implementamos então dois procedimentos de leitura de temperatura para os nós: o primeiro não recebe nenhum parâmetro e retorna o valor "crú" do sensor de temperatura, sem nenhum tratamento; o segundo recebe um inteiro que indica se o cliente deseja a temperatura em Celsius ou Fahrenheit. Neste último caso, o nó deve ser o valor do sensor e realizar uma conversão para a unidade desejada — destacando que a escala usada pelo sensor não tem relação nenhuma com as duas escalas de temperatura citadas.

A Figura 5 mostra um trecho da implementação na linguagem Céu [SantAnna 2011] (usada pelo *framework* Terra). Na linha 3, o código espera uma requisição do *gateway*, na linha 5 é solicitado ao sensor a leitura da temperatura, que é armazenada em um dos campos da mensagem que será enviada como resposta (linhas 7 e 8). Por questão de *debug*, as linhas de 11 a 20 são responsáveis por piscar os leds do nó a cada dois segundos.

Para a construção da aplicação cliente, utilizamos a linguagem Lua [Ierusalimschy et al. 1996] com a biblioteca Verse [Wild 2016] para XMPP. Lua é uma linguagem dinâmica, que permite a carga de código em tempo de execução (possibilitando a atualização do programa), e que possui também algumas características de linguagens funcionais.

A Figura 6 mostra duas versões de código para recuperar a temperatura de um nó da RSSF. Cada versão interage com uma determinada implementação de procedimento

```
1  par do
2    loop do
3      rcvMsg = await RECEIVE;
4
5      sndMsg.value = await TEMP;
6
7      emit SEND(sndMsg);
8      await SEND_DONE;
9    end
10 with
11   loop do
12     await 2s;
13
14     if info == 1 then
15       emit LED1(OFF);
16       info = 0;
17     else
18       emit LED1(ON);
19       info = 1;
20     end
21   end
22 end
```

Figura 5. Implementação da leitura da temperatura em *Céu*.

que está implantada no nós, como descrito anteriormente. Note que usamos `1.0` e `2.0` como parte da identificação dos procedimentos para diferenciá-los. De fato, não há nenhum tratamento especial por parte do *gateway*, o administrador é que deve fazer essa gerência ao cadastrar os procedimentos.

A biblioteca Verse não possui essa abstração de RPC, por isso tivemos que desenvolver uma camada de software. A função `createRPC` que desenvolvemos recebe como parâmetros o endereço XMPP do *gateway*, o identificador do nó da RSSF que irá processar efetivamente o procedimento e o identificador do procedimento cadastrado pelo administrador. Essa função retorna uma função (salva na variável `rpc`) que efetivamente irá fazer a invocação remota ao ser chamada.

Note que `rpc` recebe como parâmetros os campos para preencher o formulário solicitado pelo módulo Comandos Ad-hoc, ou seja, internamente a função `rpc` tem que fazer o casamento entre os parâmetros da função e os campos do formulário. Além disso, quando a resposta chegar, `rpc` tem que disponibilizar uma estrutura com os campos de resposta de acordo com o cadastro do procedimento.

Para a criação do RPC, empregamos o mecanismo de co-rotinas (multi-thread cooperativa) de Lua. Essa construção já foi explorado em outros trabalhos utilizando Lua [Ierusalimschy et al. 1998, Ururahy et al. 2002]. A Figura 7 mostra a implementação da função `createRPC`. Retiramos a manipulação de XML necessária para o funcionamento do XMPP a fim de facilitar a leitura do código. A função `createRPC` internamente somente cria (linhas de 3 a 31) e retorna a função `rcp`. Quando esta função é chamada pela aplicação cliente, `rpc` vai criar uma função de *callback* (linhas de 5 a 22), que será responsável por processar o formulário retornado pelo *gateway*. Na linha 27 a função `rpc`

```
local gateway = "gateway@blackbox/Server"
local node_id = 2

-- Versão 1.0
function getv1()
  local rpc = createRPC(gateway, node_id, "getTemperature::1.0")
  local resp = rpc()
  print(resp.temperature)
end

-- Versão 2.0
function getv2()
  local rpc = createRPC(gateway, node_id, "getTemperature::2.0")
  local resp = rpc({ unidade = 1 })
  print(resp.temperature)
end
```

Figura 6. Exemplo de funções em Lua para recuperar a temperatura.

dispara o procedimento remoto (passando a *callback*) e na linha 30 a *thread* (co-rotina) é suspensa (bloqueando a aplicação) esperando a liberação vinda da *callback*.

Quando o *gateway* devolver à aplicação cliente o formulário com os campos a serem preenchidos, a *callback* em um primeiro estágio vai recebê-lo, preenchê-lo e retornar ao *gateway* (linhas de 6 a 12). O *gateway* vai repassar o pedido à RSSF, e quando obtiver resposta, vai repassá-la ao cliente. Novamente a função de *callback* vai ser ativada (linhas 13 a 21), decodificando a resposta vinda do *gateway*. Na linha 20 a *callback* desbloqueia o código da aplicação, repassando a resposta. Assim, na linha 30, a função `yield` retorna a variável `resp` que finalmente será o `return` da função `rpc` para a aplicação cliente.

3.1. Teste de Validação

Utilizando as ferramentas descritas na seção anterior, o teste de validação avaliou a viabilidade da proposta, não sendo feito nenhuma medida de desempenho sobre o *gateway*, por exemplo, teste de carga.

Para o teste utilizamos dois nós TelosB[3], um sendo o sorvedouro ligado ao computador onde o *gateway* estava instalado, e o outro nó servindo como leitor de temperatura. Inicialmente, o TelosB foi carregado com a primeira versão da aplicação para RSSF (a que retorna o valor da temperatura vindo diretamente do sensor), o procedimento correspondente foi cadastrado no *gateway* e um programa Lua fez a chamada remota para obter o valor.

Utilizando as ferramentas do *framework* Terra, foi feito o *upload* da segunda versão da aplicação para RSSF para o TelosB (atualização dinâmica remota). A nova interface do procedimento foi cadastrada no *gateway* (sem a necessidade de reinicialização) e obtivemos o valor da temperatura via programa cliente em Lua invocando o novo procedimento.

Destacamos que no teste, utilizamos dois programas Lua, um para cada interface exportada pelo nó da RSSF, apenas pela simplicidade do teste, mas, como apresentado

[3]TelosB possui 10Kb de RAM, processador MSP430 de 16bits a 8MHz e rádio padrão 802.15.4.

```
function createRPC(server, mote, cmd)

  local rpc = function(params)

    local cb = function(evt)
      if evt.status == "executing" then

        -- Formulário recebido, preenchê-lo com os parâmetros do RPC.

        -- Enviar o formulário preenchido ao gateway.
        evt:next(form)

      elseif e.status == "completed" then

        -- Resposta recebida, transformar de XMPP para Lua.
        -- Variável 'resp' conterá os dados a serem retornados.

        -- Desbloquear o código da aplicação cliente,
        -- retornando a resposta vinda do gateway.
        coroutine.resume(app, resp)
      end
    end

    -- Dispara a execução do comando ad-hoc identificado por 'proc'.
    -- A callback 'cb' será acionada para tratamento do formulário.
    local proc = "wsn-proc/" .. mote .. "/" .. cmd
    conn:execute_command(server, proc, cb)

    -- Suspende a chamada até termos a resposta completa do RPC.
    return coroutine.yield()
  end

  return rpc
end
```

Figura 7. Pseudo-código da implementação do RPC em Lua.

anteriormente, Lua é uma linguagem que permite carga de código dinâmica, então não seria difícil receber um novo trecho de código via rede ou até recarregar um módulo de arquivo.

4. Trabalhos Relacionados

Hornsby *et al.* [Hornsby et al. 2009] também propuseram uma arquitetura baseada em XMPP para interação com RSSF baseada em um *gateway*. Diferente de nossa abordagem, eles não utilizaram extensões XMPP. As mensagens XMPP eram específicas para cada nó, seguindo um modelo de *request/response*, sem uso de formulários e tradução automática entre XMPP e RSSF. Eles utilizaram UPnP como serviço adicional a fim de realizar o papel de descoberta de serviço.

Em [Kirsche and Klauck 2012], os autores discutem a implementação do protocolo XMPP direto nos nós da RSSF, sem a necessidade de um *gateway*. No entanto, os próprios autores citam a restrição de memória como limitador, pois além do protocolo

XMPP existe ainda a necessidade de uso da pilha TCP/IP, o que demandaria nós com mais recursos ou deixando pouco para a aplicação em si. A vantagem do uso de um *gateway* é permitir que aplicações legadas sejam integradas à Internet ou mesmo usar menos recursos nos nós.

[Wang et al. 2012] apresentaram uma arquitetura baseada no *middleware* LooCI [Hughes et al. 2009] que possui um *proxy* dinâmico responsável integração de RSSF com computação em nuvem. Uma vantagem da proposta é que a RSSF utiliza um protocolo textual, onde as mensagens carregam o nome do campo e o seu valor. Com isso a arquitetura cria dinamicamente um componente que passa a ser um *proxy* para o nó, provendo suas informações. Nosso trabalho assume o TinyOS, que utiliza formato binário das mensagens, o que não permite essa descoberta dinâmica do formato das mensagens.

No trabalho [Kurowski et al. 2010], os autores empregam o protocolo XMPP para o monitoramento do gasto de energia de equipamentos. Eles utilizaram o conceito do protocolo REST sobre as mensagens XMPP, dando uma certa flexibilidade à comunicação, pois o formato da mensagem não possui campos rigidamente definidos, facilitando a adaptabilidade. No entanto, essa construção não segue nenhuma padronização, como por exemplo, as XEPs utilizadas em nosso trabalho.

5. Conclusão

Neste trabalho apresentamos a arquitetura de um *gateway* para acesso aos nós de uma RSSF utilizando o protocolo XMPP. Empregamos extensões padronizadas do protocolo para que o *gateway* permitisse às aplicações clientes realizarem chamadas RPC (que são repassadas à RSSF), e também um certo nível de descoberta de serviço.

Os procedimentos remotos podem ser cadastrados dinamicamente pelo administrador, podendo assim acompanhar mudanças na RSSF: se nós da RSSF são modificados para oferecer novas funcionalidade, essas mudanças podem ser incorporadas no *gateway* sem a necessidade de modificação de código e reinicialização do serviço.

Como validação da proposta, empregamos algumas ferramentas que exercitam a atualização dinamicamente da RSSF e da aplicação cliente, permitindo demonstrar a viabilidade do *gateway*.

RPC foi a abstração de comunicação utilizada neste trabalho, mas há cenários em que a RSSF emprega coletas de dados, ou seja, os nós enviam periodicamente os dados para o sorvedouro. Uma linha de evolução do *gateway* é oferecer mecanismos como *Publish-Subscribe* para acomodar esse modelo de fornecimento de dados.

A decisão de conexão da RSSF por um *gateway* é viável para casos em redes RSSF privadas ou até legadas. No entanto, vemos recentemente o conceito de IoT se tornando cada vez mais realidade, onde os nós estão acessíveis diretamente na Internet, e muitas vezes conectados a computação em nuvens, possibilitando a mineração de dados e processamento de grande volume de informação. Mas, dependendo da capacidade dos nós da rede, ou mesmo da dificuldade de atualização do software da RSSF, a interligação dos nós com a Internet e seus serviços demandariam ainda um ponto de conversão. Assim, estudar formar de acoplar o *gateway* com tecnologias em nuvens, servindo como fonte de dados, pode ser outro ponto a ser explorado. Talvez a padronização do protocolo XMPP e suas extensões possam facilitar o processo.

6. Agradecimento

Este trabalho foi parcialmente financiado pelo programa de PIBIC do CNPq.

Referências

Atzori, L., Iera, A., and Morabito, G. (2010). The Internet of Things: A survey. *Computer networks*, 54(15):2787–2805.

Bolton, F. (2001). *Pure CORBA*. Pearson Education.

Branco, A., Santánna, F., Ierusalimschy, R., Rodriguez, N., and Rossetto, S. (2015). Terra: Flexibility and safety in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 11(4):59.

Cerami, E. (2002). *Web services essentials: distributed applications with XML-RPC, SOAP, UDDI & WSDL*. O'Reilly Media, Inc.

Hildebrand, J., Millard, P., Eatmon, R., and Saint-Andre, P. (2008). XEP-0030: Service discovery. Technical report, XMPP Standards Foundation.

Hornsby, A., Belimpasakis, P., and Defee, I. (2009). Xmpp-based wireless sensor network and its integration into the extended home environment. In *13th International Symposium on Consumer Electronics*, pages 794–797. IEEE.

Hughes, D., Thoelen, K., Horré, W., Matthys, N., Cid, J., Michiels, S., Huygens, C., and Joosen, W. (2009). LooCI: a loosely-coupled component infrastructure for networked embedded systems. In *7th International Conference on Advances in Mobile Computing and Multimedia*, pages 195–203. ACM.

Ierusalimschy, R., Cerqueira, R., and Rodriguez, N. (1998). Using reflexivity to interface with CORBA. In *Conference on Computer Languages*, pages 39–46. IEEE.

Ierusalimschy, R., de Figueiredo, L. H., and Celes Filho, W. (1996). Lua-an extensible extension language. *Software: Practice & Experience*, 26(6):635–652.

Ignite Realtime (2016). Smack library.

Kaliski, R. D. (2005). *TinyOS Laboratory Development*. PhD thesis, Faculty of Cal Poly, San Luis Obispo.

Kirsche, M. and Klauck, R. (2012). Unify to bridge gaps: Bringing XMPP into the Internet of Things. In *International Conference on Pervasive Computing and Communications Workshops*, pages 455–458. IEEE.

Kurowski, K., Oleksiak, A., Witkowski, M., and Nabrzyski, J. (2010). Distributed power management and control system for sustainable computing environments. In *International Green Computing Conference*, pages 365–372. IEEE.

Levis, P., Madden, S., Polastre, J., Szewczyk, R., Whitehouse, K., a. Woo, Gay, D., Hill, J., Welsh, M., Brewer, E., et al. (2005). TinyOS: An operating system for sensor networks. In *Ambient intelligence*, pages 115–148. Springer.

Miller, M. (2015). XEP-0050: Ad-hoc commands. Technical report, XMPP Standards Foundation.

Saint-Andre, P. (2011). RFC 6120: Extensible messaging and presence protocol (XMPP): Core. RFC, Internet Engineering Task Force.

SantAnna, F. (2011). Ceu: A reactive language for wireless sensor networks. In *Proceedings of the ACM SenSys*, volume 11.

Silvestre, B., Filho, J. P., Rossetto, S., and Barcellos, V. (2013). Avaliação do desempenho de redes LLNs baseadas nas recomendações 6LoWPAN e RPL. In *XL Seminário Integrado de Software e Hardware (SEMISH)*.

Ururahy, C., Rodriguez, N., and Ierusalimschy, R. (2002). Alua: Flexibility for parallel programming. *Computer Languages, Systems & Structures*, 28(2):155–180.

Wang, W., Lee, K., and Murray, D. (2012). Integrating sensors with the cloud using dynamic proxies. In *23rd International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1466–1471. IEEE.

Wild, M. (2016). Verse – XMPP library for Lua.

Yick, J., Mukherjee, B., and Ghosal, D. (2008). Wireless sensor network survey. *Computer Networks*, 52(12):2292 – 2330.

Uma Avaliação de Desempenho de Algoritmos de Roteamento Multiobjetivos em Redes de Sensores Sem Fio

Paulo César G. de Brito[1], Vinicius N. Medeiros[2], Vinicius da C.M. Borges[1], Bruno O. Silvestre[3]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74001-970 – Goiás – GO – Brasil

{paulobrito, viniciusnunesmedeiros, vinicius, brunoos}@inf.ufg.br

Abstract. *In this article, we propose a performance evaluation in order to assess the impact of multiobjective routing algorithms on traffic performance of Wireless Sensor Networks. For this reason, the most relevant multiobjective approaches were selected. Besides, a generic routing protocol was developed for the evaluation of compared algorithms. The performance evaluation was carried out through simulations in a WSN scenario. The result analyses show that employing higher number of objectives does not always result in the best performance. However, a balanced combination of a greater number of objectives showed significant results in dense networks.*

Resumo. *Neste artigo propomos uma avaliação de desempenho de algoritmos de roteamento multiobjetivos em Redes de Sensores Sem Fio (RSSFs), analisando o impacto das diferentes combinações dos objetivos dessas abordagens para a determinação de rotas em RSSFs. Para prover essa avaliação, as abordagens mais relevantes que utilizam múltiplos objetivos na determinação das rotas foram selecionadas. Além disso, um protocolo genérico de roteamento foi desenvolvido para a avaliação de algumas dessas abordagens. A avaliação de desempenho foi realizada através de simulações em um cenário de RSSFs. A análise dos resultados mostra que empregar maior número de objetivos nem sempre resulta no melhor desempenho. No entanto, uma combinação equilibrada de um maior número de objetivos mostrou resultados significativos em redes mais densas.*

1. Introdução

A Internet apresentou nas últimas décadas um considerável crescimento, se tornando uma ferramenta essencial para o ser humano. Atualmente, não só o homem utiliza dessa tecnologia de comunicação, como também máquinas (aparelhos e dispositivos), que se comunicam entre si, fazendo medição de variados tipos de informação. A *Internet of Things* (IoT) ou Internet das Coisas vem se tornando um novo paradigma de comunicação que possibilita que dispositivos se comuniquem e monitorem o ambiente em que estão inseridos [Atzori et al. 2010].

Rede de Sensores sem Fio (RSSF) cumpre um papel importante dentro da IoT e tem ganhado cada vez mais destaque como parte da computação pervasiva/ubíqua em diversos ambientes, como indústria, cidades inteligentes, espaços inteligentes, redes de energia inteligentes, saúde, monitoramento ambiental, aplicações multimídias de tempo real, etc. [Atzori et al. 2010, Oliveira and Rodrigues 2011, Alanazi and Elleithy 2015],

RSSF pode ser composta por uma grande quantidade de nós (centenas ou milhares). Esses nós geralmente possuem forte restrição de processamento, bateria e memória. Normalmente, cada nó possui apenas uma interface de rádio (geralmente, 802.15.4 CSMA/CA) com uma taxa de transmissão fixa (250 Kbps) [Yick et al. 2008].

Uma das formas de organizar a comunicação dos nós de uma rede é no modelo *flat*, onde todos os nós desempenham o mesmo papel de sensoriamento, processamento e (re)transmissão de pacotes. As informações são repassadas (roteadas) nó a nó, até chegar a um nó *sink*, onde elas são processadas [Akyildiz et al. 2007, Niculescu 2005, Radi et al. 2011].

No roteamento, cada salto pode significar um aumento no atraso de entrega da informação para o nó *sink*. Além disso, uma quantidade maior de saltos acarreta na ativação de mais nós, o que gera mais utilização do meio de transmissão, podendo levar a mais contenção e interferência [Flushing and Di Caro 2013]. Ao reduzirmos o número de saltos, tentamos reduzir também o atraso de entrega das mensagens de rede.

O balanceamento de carga visa distribuir os fluxos de roteamento pela rede. Isso evita que um conjunto menor de nós sejam escolhidos pela maioria das rotas o que leva a uma carga maior sobre esses nós. Destacamos que os nós têm recursos limitados, assim, essa carga pode levar, por exemplo, à redução do tempo de vida desses nós, dado um consumo maior de bateria para o roteamento dos pacotes, e também à perda de pacotes e latência, dada a quantidade reduzida de memória (*buffer* cheio) para armazenamento de pacotes em rota [Kang et al. 2004].

Finalmente, temos a interferência que degrada a qualidade do enlace de comunicação. Ela pode ser causada pelo número de vizinhos que estão no alcance de comunicação de cada nó e/ou por outras redes sem fio que adotam geralmente padrões de comunicação na mesma faixa de espectro. Como os nós de uma RSSF utilizam esse meio de transmissão bastante compartilhado, e empregam apenas um rádio e um único canal de comunicação, quanto mais nós na rede, maior grau de interferência. Além disso, um maior número de vizinhos usando o mesmo canal pode causar contenção que implica uma maior chance de colisão (com necessidade de retransmissão do pacote) pois há uma maior probabilidade do canal estar ocupado. Como resultado, a interferência impacta fortemente na capacidade efetiva de transmissão [Flushing and Di Caro 2013]. Assim, as rotas que passam por nós com maior grau de interferência e contenção podem sofrer com maior atraso na entrega e/ou mais retransmissão de pacotes (impactando também no maior consumo de bateria).

Podemos notar inter-relações entre esses três critérios. O aumento do número de saltos no caminho do fluxo é um efeito comum quando uma solução de roteamento procura minimizar nós sobrecarregados [Flushing and Di Caro 2013]. No entanto, o aumento do tamanho do caminho gera uma maior ativação de nós, causando mais interferência e contenção nos vizinhos. É importante ressaltar que nem sempre caminhos mais curtos geram menos contenção/interferência nas RSSFs como um todo. Por exemplo, os caminhos mais curtos podem sofrer contenção devido à existência de vizinhos ou outras redes/dispositivos que estão na mesma área de alcance desses caminhos.

Existem vários trabalhos que tentam combinar esses objetivos ou, pelo menos, alguns deles para otimizar o roteamento em redes sem fio, especialmente em

RSSFs. A maior parte trata de um ou dois objetivos descritos acima. Wang e Zhang [Wang and Zhang 2010] apresentaram um protocolo denominado "Interference Aware Multipath Routing (IAMR)" que seleciona caminhos espacialmente disjuntos, adotando um modelo de interferência onde o alcance de interferência é o dobro do alcance de transmissão. É importante ressaltar que esse modelo pode sobre-estimar o nível de interferência.

Radi *et al.* [Radi et al. 2010] propôs "Low-Interference Energy-Efficient Multipath Routing Protocol (LIEMRO)", que leva em conta uma abordagem de balanceamento de carga de múltiplos caminhos e que também procura minimizar a interferência através do uso da métrica de roteamento ETX. Como já descrito na literatura, a métrica ETX tem limitações para capturar a interferência de uma forma mais precisa [Borges et al. 2011].

Minhas *et al.* [Minhas et al. 2009] recomendaram uma abordagem baseada em lógica *fuzzy* para problema de otimização multiobjetivo em roteamento, onde procura equilibrar dois objetivos: maximizar o tempo de vida da rede (analisando os níveis de energia atual e residual dos nós) e o atraso na transmissão entre o nó ordinário e o nó *sink* (levando em conta o número de saltos).

O roteamento com balanceamento de carga através de uma distribuição uniforme dos fluxos na rede pode prover tanto a minimização da perda de pacotes como também a maximização do tempo de vida da rede. Por essa razão, abordagens de roteamento são propostas em redes sem fios para minimizar os gargalos e o tamanho médio dos caminhos em saltos, por exemplo o algoritmo *Bottleneck, Path Length and Routing overhead* (BPR) [Mello et al. 2014].

Moghadam *et al.* [Moghadam et al. 2014] apresentaram uma heuristica denominada "Heuristic Load Distribution (HeLD)", que objetiva minimizar o *overhead* de comunicação e maximizar o tempo de vida da rede através de uma abordagem de roteamento com balanceamento de carga baseada em programação linear e o uso de um conjunto de caminhos entrelaçados. Os testes em simulação mostraram que HeLD resulta em menos *overhead* e uma rede com maior tempo de vida, entretanto, a taxa de entrega de pacotes alcançou níveis baixos e a latência nem foi apresentada. Isso pode ser parcialmente justificado pelo conjunto de caminhos entrelaçados, que pode gerar um grau elevado de interferência.

Machado *et al.* [Machado et al. 2013] apresentaram uma heurística para ambientes de IoT chamada "Routing by Energy and Link quality (REL)", que procura selecionar caminhos que minimizem o número de saltos, maximizem o tempo de vida da rede através da energia residual dos nós, e maximizem a qualidade dos enlaces nos caminhos através do uso de métricas que descrevam a interferência de uma forma mais precisa — por exemplo, *Received Signal Strength Indicator* (RSSI). No entanto, a heuristica é bastante simples e é limitada em estabelecer um equilíbrio entre os três objetivo.

Medeiros *et al.* [Medeiros et al. 2016] propõem uma heurística denominada Routing-Aware Of Path Length, Link Quality, And Traffic Load (RALL) que utiliza de métodos como soma ponderada, para poder chegar a uma solução que melhor equilibre os objetivos desejados, que são: minimizar a quantidade de saltos, quantidade de enlaces com baixa qualidade e o gargalo da rede.

Dentre os trabalhos relacionados, a maioria leva em conta dois objetivos (por

exemplo, BPR, $IAMR$ e $LIEMRO$) e poucas abordagens leva em conta três objetivos (REL e $RALL$). Além disso, foram apresentados poucos trabalhos na literatura que analisasse o impacto no trafégo de RSSFs de diferentes abordagens com objetivos distintos. Tendo em vista essa carência, esse trabalho tem como objetivo comparar os algoritmos de roteamento multiobjetivo, isto é, BPR, REL e $RALL$. Para cumprir esse propósito, foi necessário o desenvolvimento de um protocolo de roteamento que fornecesse o mapa de conectividade da rede e outras informações (por exemplo, RSSI) para os diferentes algoritmos.

Este artigo é estruturado da seguinte forma. Na seção 2 serão apresentados os algoritmos de roteamento usados neste artigo. Na seção 3 será exposto o protocolo de roteamento que desenvolvemos para fins de avaliação dos algoritmos selecionados. Na seção 4 são apresentados os resultados das simulações. A conclusão e trabalhos futuros é discutida na seção 5.

2. Algoritmos de roteamento

Nesta seção descreveremos as abordagens escolhidas para análise. Esta seção é estruturada como segue: a abordagem REL é descrita na subseção 2.1. A abordagem BPR é descrita na subseção 2.2 e a abordagem $RALL$ é descrita na subseção 2.3.

2.1. *Routing Protocol Based on Energy and Link Quality - REL*

REL [Machado et al. 2013] visa melhorar a confiabilidade e eficiência energética da rede, possibilitando a seleção de rotas baseado na qualidade do enlace, na energia residual do nó e no número de saltos. Para determinar a qualidade de comunicação dos enlaces o REL utiliza o indicador LQI (*Link Quality Indicator*) e assim propõe um contador denominado *WeakLinks* para determinar a quantidade de enlaces ruins existentes em um caminho, nem sempre o menor caminho para um destino é composto por enlaces com melhor qualidade, pode haver um caminho com maior comprimento, porém com uma melhor capacidade de transmissão.

Para prover uma maior acurácia e evitar atualizações excessivas o REL ao invés de trabalhar somente com os valores absolutos de LQI, trabalha com a média desses valores LQI para um determinado destino. Com isso, se consegue um menor atraso e menos overhead, já que existem menos pacotes de controle na rede. O mecanismo de descoberta de rotas do REL é sob demanda, com o objetivo de diminuir a sobrecarga e melhorar a escalabilidade. Os pacotes que são utilizados são o $RREQ$ (*Route Request*) e o $RREP$ (*Route Response*), que são enviados em *broadcast*, além de buscar rotas previamente estabelecidas, estes pacotes auxiliam no processo de seleção de rota, coletando informações sobre a qualidade do enlace e a energia residual dos nós, assim campos adicionais foram criados nos pacotes para armazenarem essas informações.

No processo de seleção de rotas, o REL utiliza de dois limiares para comparar possíveis rotas, que são um limiar para tamanho máximo de saltos e outro limiar para a energia residual. Quando o valor de energia gasto de um nó ultrapassar o limiar de energia residual, um processo para o estabelecimento de novas rotas é iniciado na rede. E já o limiar de tamanho máximo de saltos é utilizado como critério na seleção de rotas, excluindo rotas que ultrapassem este valor. No entanto, essa abordagem falha em obter uma combinação entre diferentes objetivos, trabalhando com os mesmos de forma independente. Além disso, não leva em conta o balanceamento de carga.

2.2. *Bootleneck, Path length and Routing overhead -BPR*

BPR [Mello et al. 2014] é uma heurística que tem como finalidade obter uma melhora no balanceamento de carga e no comprimento dos caminhos. Ele tem como prioridade minimizar o gargalo da rede e em seguida minimizar o comprimento dos caminhos estabelecidos para os fluxos da rede.

O algoritmo BPR é executado quando um novo fluxo é iniciado. Em outras palavras, ele opera quando são iniciados os fluxos do gargalo (aresta com maior número de fluxos da rede). Essa escolha é feita com a finalidade de evitar a frequente execução da solução de roteamento para todos os fluxos da rede, pois isto pode causar um alto nível de oscilação do roteamento devido a ordem na qual os fluxos são roteados.

O algoritmo utiliza o fluxo a ser adicionado, o conjunto de arestas do grafo que representa a rede e um conjunto de caminhos candidatos para determinar a rota do fluxo. Sendo o conjunto de caminhos candidatos formados pelos possíveis caminhos entre o Internet gateway (isto é, $sink$) e os roteadores sem fio ordenados em ordem crescente de comprimento de caminho e limitados por um fator de esticamento (número máximo de saltos). A saída do algoritmo é o conjunto de arestas com seus valores atualizados, isto é, representando a contagem de fluxos que passam por cada aresta.

Um ponto a se observar nessa abordagem, é que ela não leva em conta qualidade do enlace de comunicação, o que de certa forma é ruim, já que ao gerar um balanceamento de carga, caminhos que apresentam baixa qualidade de transmissão podem ser usados ao evitar nós sobrecarregados, o que aumenta a taxa de perdas e atraso.

2.3. *Routing-Aware Of Path Length, Link Quality, And Traffic Load - RALL*

$RALL$ [Medeiros et al. 2016] apresenta uma proposta para determinar o roteamento dos pacotes em RSSFs utilizando a combinação de três objetivos: minimizar a quantidade de saltos, a quantidade de enlaces com baixa qualidade e o gargalo da rede. O algoritmo $RALL$ utiliza algumas técnicas para normalizar os dados e possibilitar a seleção de uma solução pertencente ao conjunto pareto-ótimo. O $RALL$ primeiramente utiliza o método das somas ponderadas [Marler and Arora 2009] para combinar os objetivos de quantidade de saltos e qualidade dos enlaces em uma nova função objetivo resultante que representa um *trade-off* entre os dois objetivos. A utilização da função objetivo resultante para a determinação de rotas gera caminhos que satisfaçam os objetivos usados em sua criação, dependendo exclusivamente dos pesos utilizados na soma.

O desenvolvimento de um algoritmo que consiga determinar uma solução pertencente ao conjunto Pareto-ótimo (conjunto formado por todas as soluções não-dominadas, dentre as soluções factíveis) para o problema de otimização multiobjetivo possui grande complexidade computacional. Tal dificuldade é criada principalmente pelo objetivo referente à minimização do gargalo por apresentar uma representação matemática distinta dos outros objetivos abordados, minimizar a quantidade de saltos e minimizar o uso de enlaces com baixa qualidade. A determinação do gargalo em uma rede exige o cálculo de todas as rotas para determinar o valor da função.

Por essa razão, o $RALL$ determina uma solução para o problema de otimização utilizando o terceiro objetivo de forma incremental, ou seja, adicionando fluxo por fluxo sem a necessidade de ter previamente todas as rotas definidas. Em outras palavras, para

selecionar rotas que consigam minimizar também o gargalo da rede a cada novo caminho determinado, a função objetivo resultante tem seus coeficientes, que representam as arestas utilizadas por cada rota, incrementados/atualizados.

O $RALL$ não leva em conta um objetivo importante em RSSFs, que é maximizar eficiência energética (por exemplo, energia residual dos nós, tempo de vida da rede).

3. Protocolo de roteamento

Desenvolvemos um protocolo de roteamento visando possibilitar a avaliação de algumas abordagens descritas nesse artigo, isto é BPR e $RALL$. A necessidade do desenvolvimento surgiu após detectarmos que os protocolos existentes não estavam satisfazendo os requisitos necessários para a execução dos algoritmos, como incluir informações específicas em pacotes de controle utilizados em cada algoritmo avaliado e também devido ao $overhead$ demasiado que estes protocolos produzem (exemplo: OLSR) e que não agregam em nosso estudo. Esse protocolo foi implementado na linguagem C++ e utilizando módulos nativos do simulador OMNeT++, $plugin$ Castalia. O protocolo trabalha com pacotes de controle que possibilitam a descoberta da rede, a disseminação de informações e a descoberta de rede desconexa.

Os pacotes de controle envolvidos nesse protocolo são diferenciados em sua estrutura para atenderem às necessidades de cada mensagem de controle citadas acima e são divididos em três:

- *Descoberta de vizinhos:* Como mostrado na figura 1, este é um pacote genérico utilizado para identificar os vizinhos de um nó, a partir da transmissão desse pacote disseminamos duas informações: (i) valor de LQI (*Link Quality Indicator*) e (ii) nó de origem do pacote.

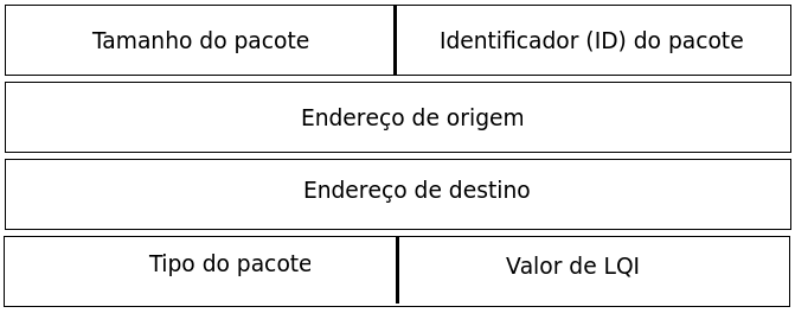


Figura 1. Cabeçalho do pacote de descoberta.

- *Disseminação:* Este pacote tem sua estrutura modificada dependendo do tipo de disseminação, que pode ser: (1) Dos nós para o $sink$ ou (2) do $sink$ para os nós, na figura 2 ilustramos esses dois casos. Quando a disseminação é do tipo (1) o pacote leva consigo informações a respeito dos nós vizinhos com seus respectivos valores de LQI do remetente. Quando o pacote de disseminação é do tipo (2) o pacote leva uma tabela de roteamento, que servirá para os nós se orientarem sobre qual é o próximo salto para encaminhar os pacotes de outros nós.
- *Notificação:* O pacote de notificação tem a finalidade de relatar a inatividade de algum nó na rede, este pacote conta com os campos de nó monitor, isto é, o nó que monitora seu vizinho e o nó que está possivelmente inativo, conforme ilustrado na

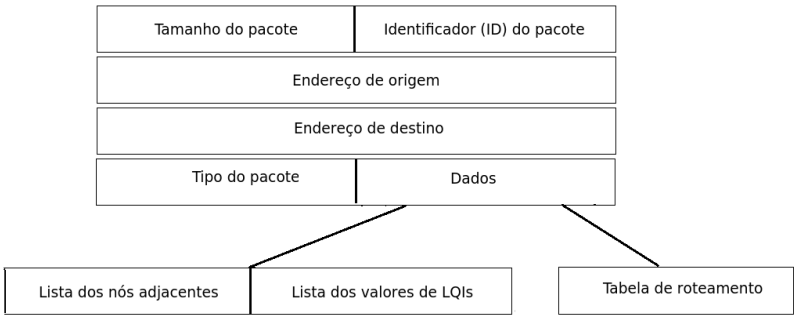


Figura 2. Cabeçalho do pacote de disseminação.

figura 3. Assim com posse dessas duas informações, o mapa de conectividade é modificado e então o algoritmo de roteamento é re-executado no *sink*, iniciando o processo que envolve recalcular as rotas sem utilizar o nó inativo.

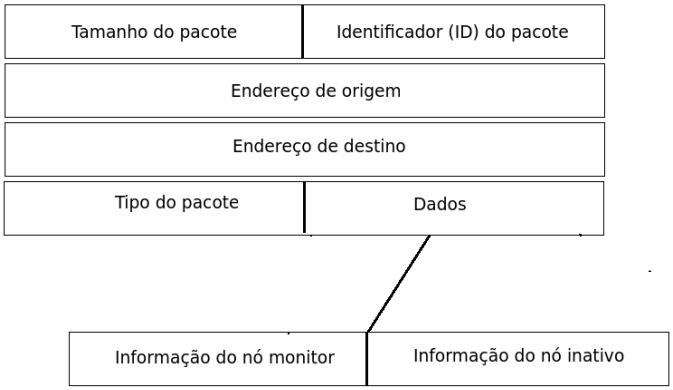


Figura 3. Cabeçalho do pacote de notificação.

Inicialmente o protocolo realiza a descoberta dos nós vizinhos de cada nó da topologia, conforme é ilustrado na figura 4, para isso cada nó envia pacotes de descoberta de vizinhos em modo de comunicação *broadcast*, os nós vizinhos que recebem esses pacotes de descoberta, armazenam esses dados (nó remetente e LQI) em uma estrutura específica. Ao mesmo tempo que estes pacotes são coletados para o reconhecimento de nós vizinhos, é realizado o cálculo da média de LQI (*Link Quality Indicator*) entre esses pacotes, ou seja, para cada par de vizinhos, há uma média de LQI que poderá ser usada pelos algoritmos.

Após a descoberta de vizinhos, ilustrada na figura 4, os nós começam a enviar pacotes de disseminação do tipo (1) com as informações de nós adjacentes e suas médias de LQI para o $sink$ em *broadcast*, de posse dessas informações o algoritmo de roteamento executa e calcula a rota para cada nó da topologia.

Durante a simulação para suavizar o *overhead* na rede e um atraso na entrega de pacotes, utilizamos de um limiar para o envio de pacotes de disseminação do tipo (1) e assim reduzir o *overhead* na rede, que envolve a média de valores LQI, assim, se

houver uma mudança muito discrepante no LQI um novo pacote de disseminação (1) com seus valores de LQI é enviado ao $sink$ para que este possa calcular uma nova tabela de roteamento.

Após a execução do algoritmo de roteamento, há a disseminação dos pacotes gerados no $sink$ que contém as tabelas de roteamento para cada nó da topologia, como exemplificado na figura 4, essa disseminação ocorre em modo de comunicação *broadcast*. Ao receber a sua tabela de rotas do $sink$, cada nó identifica qual é o seu nó destino para transmissão de pacotes de dados. Após essa identificação, inicia-se um monitoramento de atividade desse nó destino, atividade esta que é percebida por chegada de pacotes de descoberta de vizinhos. Caso após um tempo estabelecido não houver nenhum indício de atividade, um pacote de notificação é enviado ao nó $sink$.

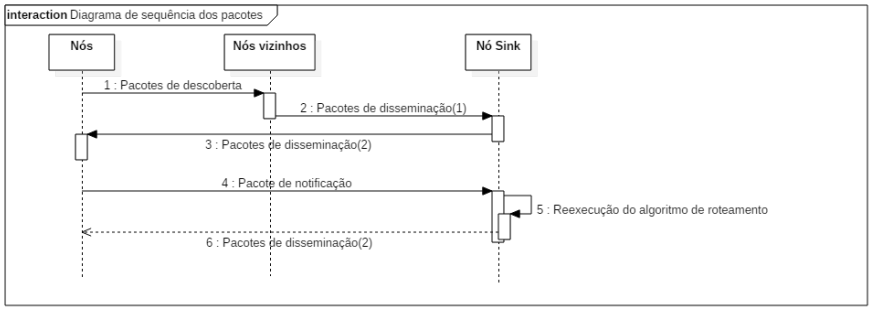


Figura 4. Diagrama de sequência dos pacotes.

4. Resultados das simulações

Nessa seção, nós conduzimos um estudo de simulação para analizar o desempenho das abordagens comparadas. A seção é estruturada como segue: o cenário de configuração é esboçado na subseção 4.1. As avaliações e os resultados das simulações são discutidas na subseção 4.2.

4.1. Configuração do ambiente

Nós empregamos o módulo Castalia [Boulis 2011] do simulador OMNeT++ [Varga 2001], que é bastante usado para avaliar RSSFs. A tabela 1 mostra os parâmetros gerais usados na simulação e a configuração dos nós baseado em [Moghadam et al. 2014]. Uma rede com um único nó $sink$ e vários nós sensores também são incluídos na rede simulada. O nó $sink$ recebe dados dos nós sensores e serve como uma ponte de conexão com a rede externa (por exemplo, Internet). Os nós sensores mandam os dados coletados e repassam pacotes através de uma única rota para o $sink$. Assim, há um único fluxo para cada nó sensor. Todos os nós são estáticos e usam enlaces assimétricos.

As simulações foram realizadas em cenários onde o número de nós variou de 10 até 50 nós. Para cada conjunto de nós, dez topologias e quatro sementes aleatórias foram geradas, resultando em um total de quarenta execuções para cada conjunto de nós. As posições dos nós na topologia foi gerada através de uma distribuição normal, em uma área limitada, para garantir que nenhum dos nós ficaria isolado na rede (ex: fora de alcance de qualquer outro nó).

Tabela 1. Parâmetros da simulação

Configuração da Rede	
Taxa de geração de pacotes de dados	5 pacotes/minuto
Taxa de geração de pacotes de descoberta	1 pacote/30 segundos
Tempo gasto na fase de pacotes de descoberta	350 segundos
Tempo gasto na fase de disseminação	100 segundos
Modelo de geração de tráfego	Poisson
Área da topologia	50x50
Modelo de interferência	Aditivo [1]
Configuração dos Nós	
Energia inicial	100 J
Energia TX	20 mJ/pkt
Energia RX	10 mJ/pkt
Camada MAC	T-MAC
Modelo de rádio	CC2420
Poder de transmissão	-10dBm
Tamanho do pacote	127 bytes

4.2. Resultados

O objetivos da simulação é avaliar o impacto das diferentes combinações dos objetivos nos distintos algoritmos de roteamento multiobjetivos analisados. Por essa razão, uma avaliação de desempenho com vários paramêtros de desempenho de tráfego (isto é, vazão média por fluxo, latência, tempo de vida da rede, overhead, tamanho médio do caminho) que possibilite assim uma comparação profunda e detalhada entre os algoritmos analisados. Um terceiro algoritmo de roteamento foi escolhido para auxiliar na avaliação: $BALANCED - LQI$ que é baseado em uma heurística muito similar ao $RALL$, empregando apenas balanceamento de carga e funções objetivas de qualidade do enlace, e não utiliza a combinação de funções usando a soma de pesos. Esse algoritmo considera objetivos muito similares a $IAMR$ e $LIEMRO$, essas abordagens consideram somente os critérios de balanceamento de carga e qualidade do enlace. Portanto, é uma abordagem aproximada a alguns relevantes trabalhos relacionados. É importante ressaltar também que BPR, $RALL$ e $BALANCED - LQI$ empregam o protocolo de roteamento desenvolvido e apresentado nesse artigo, enquanto que REL usa um protocolo próprio.

As figuras 5 e 6 mostram a taxa de perda de pacotes e a vazão média por fluxo, respectivamente. Essas figuras ilustram que o algoritmo $RALL$ consegue alcançar melhor desempenho de trafego, especialmente em um rede de alta densidade (50 nós). Isto pode ser explicado pelo fato que o $RALL$ leva em conta a combinação mais equilibrada de três objetivos que resultam na menor perda e na maior vazão onde há uma rede densa. Consequentemente, a taxa de dados entregue com sucesso aumenta. A abordagem REL também empregar três critérios, no entanto resulta na maior perda entre um número médio e alto de nós. Isto pode ser explicado pelo fato de que esta abordagem não emprega balanceamento de carga e falha em estabelecer um equilíbrio entre o número de saltos e

[1]Interferência aditiva: usa os valores SINR (Signal to Interference plus Noise Ratio) dos pacotes, o receptor recebe o maior sinal entre duas transmissões (se este for forte o suficiente).

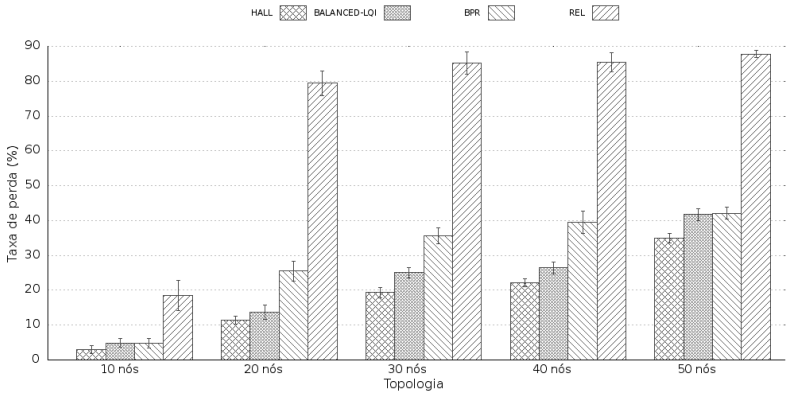


Figura 5. Taxa de perda variando o número de nós.

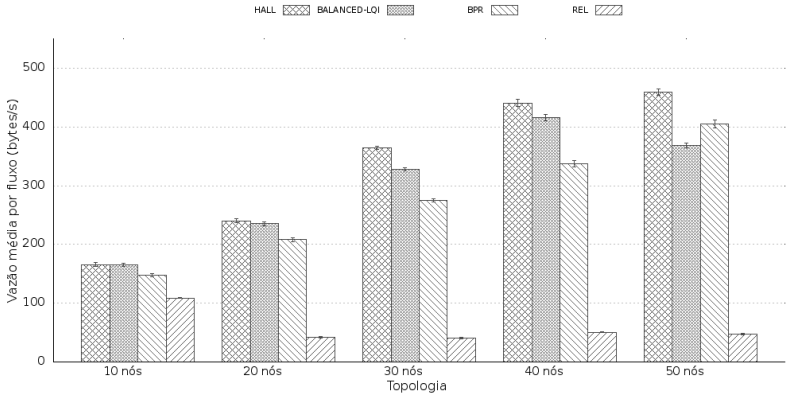
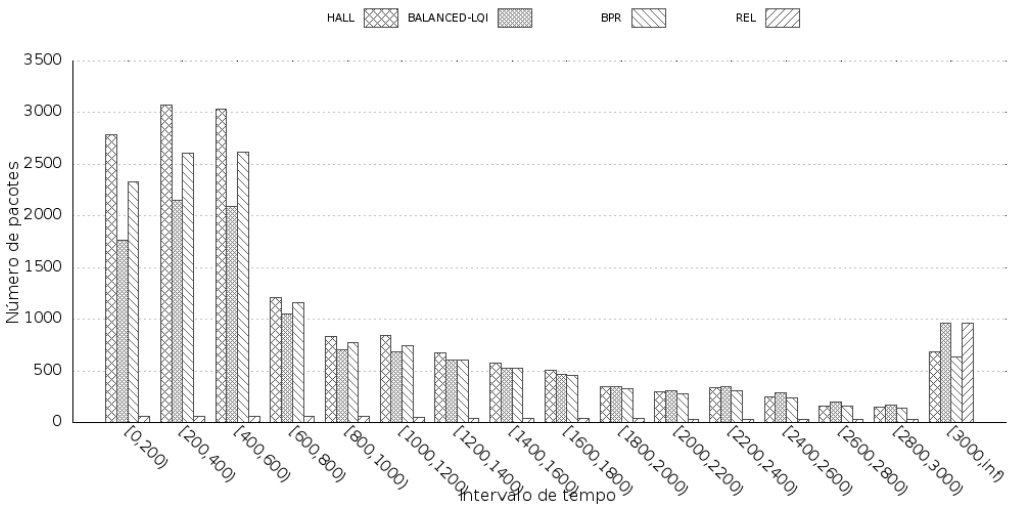


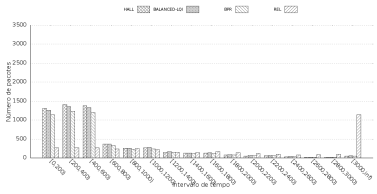
Figura 6. Vazão média.

os enlaces de baixa qualidade. $BALANCED - LQI$ e BPR combinam dois critérios, $BALANCED - LQI$ conquista menor perda de pacotes do que o BPR na maioria das topologias (10, 20, 30 e 40 nós), porque inclui ***LQI*** que tem significante influência na perda e na vazão, exceto em uma rede de alta densidade (ex: 50 nós), onde há um alto nível de interferência e contenção. Como resultado, uma taxa de perda de pacotes levemente menor e uma maior vazão é alcançada quando o BPR é utilizado, uma vez que o número de enlaces com alta qualidade é bem pequeno e a redução do tamanho do caminho minimiza os níveis de contenção e interferência.

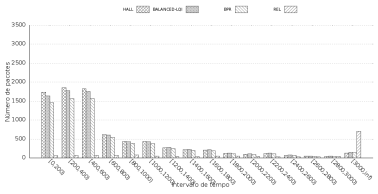
A figura 7 mostra a latência média na entrega de dados nos cenários de simulação de rede. Vale a pena notar que em todos os cenários com diferentes números de nós, a maioria dos pacotes foram entregues dentro da faixa de $[0, 600]$ ms. A grande diferença na latência entre as abordagens pode ser encontrada quando o número de nós aumenta. A abordagem $RALL$ alcança o maior número de pacotes em um pequeno intervalo de tempo $[0, 600]$ ms. Novamente, $BALANCED - LQI$ e BPR mostram uma alternância similar a encontrada na vazão e na perda, BPR resulta em um número maior de pacotes no intervalo de tempo de $[0, 600]$ ms quando o número de nós é maior. Os resultados de latência de $RALL$, $BALANCED - LQI$ e BPR podem ser parcialmente justificados pela análise do tamanho médio de caminho gerado por cada abordagem, mostrados na figura 8). Apesar do tamanho medio do caminho ter o menor valor quando REL é empregado, ele curiosamente resulta em um número bem maior de pacotes em intervalos de



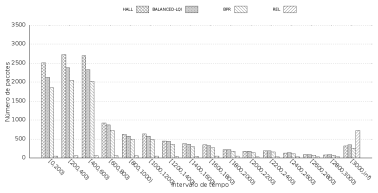
(a) 50 nós



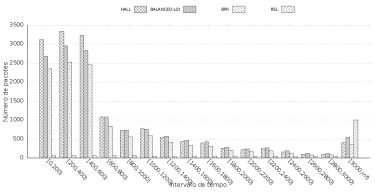
(b) 10 nós



(c) 20 nós



(d) 30 nós



(e) 40 nós

Figura 7. Latência para 10, 20, 30, 40, e 50 nós.

tempo maiores, evidenciando que nem sempre caminhos mais curtos obtêm as menores latências devido a alto níveis de interferência ou carga de trafégo nos nós que fazem parte desses caminhos mais curtos.

Nós também analizamos o tempo de vida médio da rede, que foi calculado do começo da simulação até o primeiro nó ativo (vivo) ficar sem rota para o *sink* devido a morte de todos seus nós vizinhos, ou seja completamente sem comunicação (ilhado) [Wenjing Guo and Wei Zhang 2014]. A figura 9 ilustra o tempo de vida. Deve ser ressaltado que as abordagens que resultam em uma taxa de maior pacotes entregues em uma menor latência mostram um maior tempo de vida. É importante ressaltar que abordagens que resultam em menores quantidades de pacotes de controle (isto é, overhead) consequentemente resultarão em maior tempo de vida para rede, conforme mostrado na tabela 2.

Além disso, menores tamanho médio de caminho também ajudam a aumentar o tempo de vida, especialmente em redes densas e sobrecarregadas. Vale ressaltar que

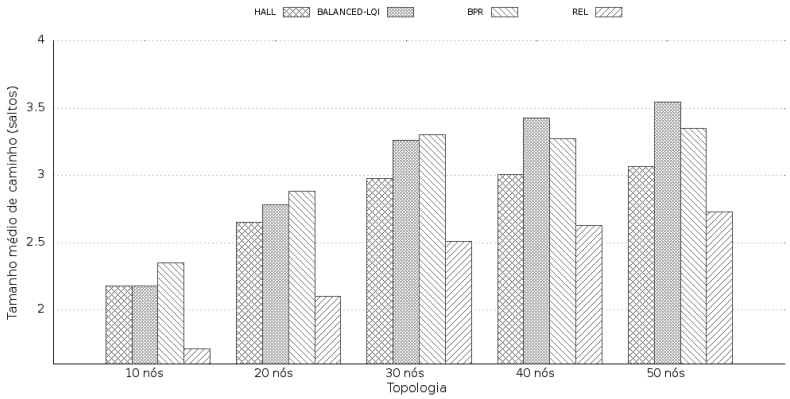


Figura 8. Tamanho médio de caminho.

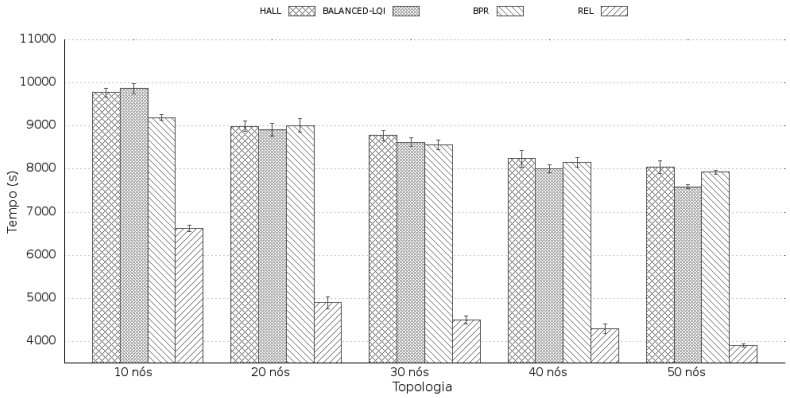


Figura 9. Tempo de vida médio.

pequenos caminhos diminuem o número de links ativos na rede, assim também reduz o consumo de energia na transmissão e recepção de pacotes de dados e de controle, uma vez que todos esse tipos de pacotes demandam um gasto para enviá-los e recebê-los. Isto também explica porque o tempo de vida da rede conquista os maiores tempo seguindo os menores caminhos e abordagens que resultem em menos overhead (isto e, ativando menor número de enlaces). Para essas razões, $RALL$, $BALANCED - LQI$ e BPR apresentam tempo de vida similares, onde $RALL$ apresenta um tempo de vida levemente maior em uma configuração de rede mais densa. REL possui o mais baixo tempo de vida justificado pelo *overhead* demasiado,que é causado pelas mudanças na topologia da rede, e na altíssima taxa de perdas de pacotes que o mesmo sofre.

Tabela 2. Quantidade de Pacotes de Controle do Protocolo de Roteamento (*Overhead*).

	RALL	BPR	BALANCED-LQI	REL
10 nós	9368.02	10825.56	10607.68	88903.34
20 nós	19234.88	21291.04	21580.46	161420.56
30 nós	36206.22	34708.28	41342.72	232798.74
40 nós	57391.16	59871.67	65244.76	302350.47
50 nós	78907.86	80366.21	94863.56	377114.08

5. Conclusões e trabalhos futuros

Nós propomos uma avaliação de desempenho de algoritmos de roteamento multiobjetivo para RSSFs nesse artigo e a implementação de um protocolo de roteamento para avaliar os algoritmos analisados. Os resultados mostraram que o algoritmo $RALL$ leva à uma significante melhora na perda de pacotes, latência, vazão média, tamanho médio de caminho, e tempo de vida da rede. As abordagens BPR e $BALANCED - LQI$ resultam em diferentes desempenhos do tráfego conforme o nível de interferência e carga na rede. REL apresentou o pior resultados dentre as abordagens analisadas. Portanto, empregar maior número de objetivos nem sempre resulta no melhor desempenho. Por exemplo, vemos algumas poucas situações que os algoritmos com dois objetivos se assemelham com um algoritmo com três objetivos. No entanto, uma combinação equilibrada de três objetivos mostrou resultados expressivos em redes mais densas.

Como trabalho futuro, pretendemos desenvolver uma técnica para a generalização das coletas de informações sobre a rede utilizando o protocolo.

Referências

Akyildiz, I. F., Melodia, T., and Chowdhury, K. R. (2007). A survey on wireless multimedia sensor networks. *Comput. Netw.*, 51(4):921–960.

Alanazi, A. and Elleithy, K. (2015). Real-time QoS routing protocols in wireless multimedia sensor networks: Study and analysis. *Sensors*, 15(9):22209.

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, 54(15):2787 – 2805.

Borges, V. C. M., Curado, M., and Monteiro, E. (2011). Cross-Layer Routing Metrics for Mesh Networks: Current Status and Research Directions. *Computer Communications*, 34(6):681 – 703.

Boulis, A. (2011). Castalia: A simulator for wireless sensor networks and body area networks. *NICTA: National ICT Australia*.

Flushing, E. F. and Di Caro, G. A. (2013). A flow-based optimization model for throughput-oriented relay node placement in wireless sensor networks. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pages 632–639, New York, NY, USA. ACM.

Kang, J., Zhang, Y., and Nath, B. (2004). End-to-end channel capacity measurement for congestion control in sensor networks. In *Proceedings of the 2nd International Workshop on Sensor and Actor Network Protocols and Applications (SANPA'04)*.

Machado, K., Rosário, D., Cerqueira, E., Loureiro, A. A. F., Neto, A., and de Souza, J. N. (2013). A routing protocol based on energy and link quality for internet of things applications. *Sensors*, 13(2):1942.

Marler, R. T. and Arora, J. S. (2009). The weighted sum method for multi-objective optimization: new insights. *Structural and Multidisciplinary Optimization*, 41(6):853–862.

Medeiros, V. N., Santana, D. V., Silvestre, B., and Borges, V. C. M. (2016). Rall - routing-aware of path length, link quality, and traffic load for wireless sensor networks. *arXiv*.

Mello, M. O. M. C., Borges, V. C. M., de L. Pinto, L., and Cardoso, K. V. (2014). Load balancing routing for path length and overhead controlling in wireless mesh networks. In *IEEE Symposium on Computers and Communications, ISCC 2014, Funchal, Madeira, Portugal, June 23-26*, pages 1–6.

Minhas, M. R., Gopalakrishnan, S., and Leung, V. C. M. (2009). Multiobjective routing for simultaneously optimizing system lifetime and source-to-sink delay in wireless sensor networks. In *Proceedings of the 2009 29th IEEE International Conference on Distributed Computing Systems Workshops*, ICDCSW '09, pages 123–129, Washington, DC, USA. IEEE Computer Society.

Moghadam, M., Taheri, H., and Karrari, M. (2014). Minimum cost load balanced multipath routing protocol for low power and lossy networks. *Wireless Networks*, 20(8):2469–2479.

Niculescu, D. (2005). Communication paradigms for sensor networks. *IEEE Communications Magazine*, 43(3):116–122.

Oliveira, L. and Rodrigues, J. (2011). Wireless sensor networks: a survey on environmental monitoring. *Journal of Communications*, 6(2).

Radi, M., Dezfouli, B., Bakar, K. A., Razak, S. A., and Nematbakhsh, M. A. (2011). Interference-aware multipath routing protocol for QoS improvement in event-driven wireless sensor networks. *Tsinghua Science & Technology*, 16(5):475 – 490.

Radi, M., Dezfouli, B., Razak, S. A., and Bakar, K. A. (2010). Liemro: A low-interference energy-efficient multipath routing protocol for improving QoS in event-based wireless sensor networks. In *Sensor Technologies and Applications (SENSORCOMM)*, pages 551–557.

Varga, A. (2001). The OMNeT++ discrete event simulation system. In *Proceedings of the European simulation multiconference (ESM'2001)*, volume 9, page 65.

Wang, Z. and Zhang, J. (2010). Interference aware multipath routing protocol for wireless sensor networks. In *IEEE GLOBECOM Workshops*, pages 1696–1700.

Wenjing Guo and Wei Zhang (2014). A survey on intelligent routing protocols in wireless sensor networks. *Journal of Network and Computer Applications (Elsevier)*, 38:185 – 201.

Yick, J., Mukherjee, B., and Ghosal, D. (2008). Wireless sensor network survey. *Computer Networks*, 52(12):2292 – 2330.

PNDVI: Um novo Algoritmo para Processamento Paralelo do índice de vegetação NDVI

Sávio S. Teles de Oliveira[1], **Wellington Santos Martins**[1], **Everton Lima Aleixo**[1], **Mateus Ferreira e Freitas**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Bloco IMF I, Campus II - Samambaia – Goiânia – GO – Brasil

savio.teles,everton.lima@gogeo.io, wellington,mateusferreirafreitas@inf.ufg.br

Abstract. *Spectral vegetation indices have been widely used to monitor vegetation cover of the Earth. Among these indices, the NDVI (Normalized Difference Vegetation Index) is the most widespread because it allows assessing the change in the green area in a certain period of time. With the volume of remote sensing data it is necessary more efficient computational solutions to this processing. In this direction, there are proposals that make intensive use of parallelism in MultiCore (CPU) and/or ManyCore (GPU) using low cost architectures. This paper presents an efficient, and low cost, solution for processing the NDVI index in large parallel satellite images on the CPU and GPU handling with the machine's memory limit. The proposed algorithm, called PNDVI (Parallel NDVI), uses a strategy of blocks to process large images, in addition to working with massive parallelism of ManyCore accelerators. The implementation on the GPU of the algorithm is shown to be nine times faster when compared with a sequential implementation and allows the processing of large images using a modest computing platform.*

Resumo. *Índices espectrais de vegetação têm sido largamente utilizados para monitorar a cobertura vegetal da Terra. Entre estes índices, o NDVI (Normalized Difference Vegetation Index) é o mais difundido, pois permite avaliar a variação da área verde em certo período de tempo. Com o aumento do volume de dados de sensoriamento remoto faz-se necessário soluções computacionais mais eficientes para este processamento. Nesta direção, estão incluídas propostas que fazem uso intensivo de paralelismo em arquiteturas MultiCore (CPU) e/ou ManyCore (GPU) de mais baixo custo. Este trabalho apresenta uma solução eficiente, e de baixo custo, para o processamento do índice NDVI em grandes imagens de satélite de forma paralela na CPU e na GPU lidando com os limites de memória da máquina. O algoritmo proposto, denominado PNDVI (Parallel NDVI), utiliza uma estratégia de blocos para processar grandes imagens, além de trabalhar com paralelismo massivo dos aceleradores ManyCore. A implementação do algoritmo na GPU se mostrou nove vezes mais rápida quando comparada com sua implementação sequencial, além de permitir o processamento de imagens grandes usando uma plataforma computacional modesta.*

1. Introdução

Índices espectrais de vegetação têm sido largamente utilizados para monitorar a cobertura vegetal da Terra em escalas global e/ou local [Miura et al. 2001]. Tais índices são

combinações de dados espectrais de duas ou mais bandas, selecionadas com o objetivo de sintetizar e melhorar a relação desses dados com os parâmetros biofísicos da vegetação. Entre estes índices, o *Normalized Difference Vegetation Index* (NDVI) é o mais difundido [Cohen et al. 2003] e permite avaliar a variação da área verde em certo período de tempo.

O volume de dados de sensoriamento remoto continuamente adquiridos por satélites e sensores aerotransportados tem aumentado rapidamente [Plaza and Chang 2007, Liu et al. 2011]. Esta proliferação de dados trouxe vários desafios as aplicações de sensoriamento remoto, mais especificamente as que necessitam do cálculo do NDVI, para armazenar e processar essa quantidade de dados.

Em um ambiente computacional com vários núcleos de processamento, o cálculo do NDVI é altamente paralelizável, já que pode ser realizado sobre cada pixel da imagem de forma independente por um núcleo de processamento. Por isso, a utilização dos vários núcleos de processamento da CPU têm sido bastante utilizado neste cálculo [Zhu et al. 2012], de forma que cada núcleo da CPU fique responsável pelo cálculo do NDVI de um conjunto de *pixels*.

Para aumentar a eficiência do cálculo do NDVI, passou-se a utilizar as unidades de processamento gráficos (GPU) [Chen et al. 2013], devido ao seu alto poder de processamento e grande quantidade de núcleos. Além disso, a arquitetura da GPU se aproveita do alto grau de paralelismo do cálculo do NDVI, onde não há necessidade de sincronismo de *threads*.

Com a evolução da tecnologia de coleta de imagens de satélite, estas imagens coletadas possuem resolução e, consequentemente, tamanho cada vez maior. Estas imagens são, em alguns casos, maiores que a memória RAM disponível na máquina, o que traz um desafio de processar estas imagens de forma eficiente. Esse desafio aumenta com o uso da GPU cuja memória é geralmente menor que a memória da CPU.

Neste contexto, este trabalho apresenta um novo algoritmo para processamento paralelo do índice NDVI utilizando uma solução paralela, denominada PNDVI (Parallel NDVI). O algoritmo foi projetado para arquiteturas Multi e *ManyCore* com o objetivo de trabalhar com imagens de satélite de qualquer tamanho, tornando viável processá-las com as restrições de memória da CPU e da GPU.

O artigo está organizado como se segue. A Seção 2 apresenta os trabalhos encontrados na literatura para processamento paralelo de imagens de satélite utilizando arquiteturas paralelas. A Seção 3 apresenta uma visão geral do algoritmo de processamento do índice NDVI na GPU e na CPU. O novo algoritmo proposto neste trabalho, PNDVI, é descrito na Seção 4 e a Seção 5 valida o algoritmo PNDVI com os experimentos e discussões sobre o desempenho do algoritmo. A Seção 6 apresenta as conclusões do artigo.

2. Trabalhos Correlatos

O processamento paralelo utilizando muitos núcleos da CPU se tornou uma solução promissora para acelerar o processamento de imagens de satélite. Por isso, grandes esforços têm sido despendidos no processamento de imagens de satélite [Plaza and Chang 2007, Ma et al. 2009], nas áreas de processamento de imagens multiespectrais [Plaza et al. 2010], mosaicos [Wang et al. 2010] e sensoriamento remoto orientado a desastre [Qu et al. 2010].

O algoritmo em [Saribekyan 2013] propõe uma estratégia de particionamento da imagem de entrada para balancear blocos de imagem entre os núcleos da CPU. Uma estratégia de paralelização do cálculo do NDVI é proposta em [Shrestha et al. 2006] utilizando uma técnica denominada *Temporal Map Algebra* (TMA). O trabalho de [ACHHAB et al. 2010] apresenta um estudo do tempo de execução de algoritmos de sensoriamento remoto, tal como o NDVI, com o uso de *Multithreading*.

Com a alta capacidade de processamento, tamanho pequeno e baixo consumo de energia, as GPUs estão conduzindo uma nova direção no processamento paralelo de imagens de satélite. Thomas et al. [Thomas et al. 2009] implementou um sistema de processamento de imagens de câmera de sensores aerotransportados. Gupta et al. [Gupta et al. 2011] propõe uma abordagem de casamento de imagens utilizando a GPU. Alvarez et al. [Alvarez-Cedillo et al. 2014] apresenta uma implementação do NDVI utilizando a biblioteca *Thrust* de CUDA. Em [Qu et al. 2013] é implementada a técnica SAM (*Spectral Angle Mapper*) utilizando a GPU, realizando uma comparação de performance com a plataforma MATLAB. Estes artigos, entretanto, não propõem nenhuma solução para o cálculo do NDVI em imagens de satélite.

Em [Christophe et al. 2011] é apresentado um *framework* que permite que algoritmos de processamento de imagens de satélite sejam implementados na GPU. Este trabalho, entretanto, não apresentou nenhuma otimização do processamento do algoritmo NDVI na GPU e, por isso, o tempo de execução na GPU foi pior do que na CPU. Em [Chemin 2010] é apresentada uma descrição completa da implementação do algoritmo NDVI na GPU utilizando a linguagem CUDA da NVIDIA e a biblioteca GDAL para leitura dos dados da imagem. Nictua et al. [Nictua and ALDEA] calcula os índices NDVI e NDWI de imagens de satélite utilizando GPU para identificar o aumento da área urbana e a redução da área de vegetação de uma região na Romênia. Estes trabalhos não projetaram os algoritmos para tratar de imagens que cujo tamanho seja maior que a memória disponível na memória da GPU. Desta forma, é inviável realizar o processamento de grandes imagens de satélite na GPU.

Para conseguir processar imagens cujo tamanho é maior que a memória disponível na GPU, o trabalho de [Juan and Jianchao 2013] carrega toda a imagem para a memória RAM da máquina e gera diversos blocos da imagem com tamanho menor que a memória da GPU. O algoritmo de cálculo do NDVI em [Juan and Jianchao 2013] apresentou um *speedup* de até 8 vezes. Em [Liu et al. 2014] são propostas estratégias de otimização utilizando uma lista na memória RAM para armazenar blocos da imagem de entrada para serem processados na GPU. Entretanto, em [Juan and Jianchao 2013] e [Liu et al. 2014] não existe nenhum tratamento para lidar com grandes imagens de satélite que não cabem na memória da CPU.

Neste sentido, este trabalho apresenta uma solução completa para realizar o cálculo do NDVI de forma paralela na CPU ou na GPU dependendo dos recursos de *hardware* disponíveis considerando a existência de imagens maiores que os limites de memória.

3. Processamento paralelo do NDVI

O índice de vegetação da diferença normalizada (NDVI) é calculado pela diferença entre as bandas do Infra Vermelho Próximo e do Vermelho, normalizada pela soma das mesmas

bandas, de acordo com a Equação 1. Os valores obtidos do NDVI estão contidos em uma mesma escala de valores entre -1 e +1.

$$NDVI = \frac{IVP - V}{IVP + V} \tag{1}$$

sendo:

NDVI - valor do índice de vegetação da diferença normalizada

IVP - valor da reflectância na faixa do infravermelho próximo,

V - valor da reflectância na faixa do vermelho.

O cálculo do índice NDVI é realizado utilizando as bandas 3 e 4 de uma imagem de satélite. A banda 3 da imagem contém os valores de V e a banda 4 os valores de IVP na Equação 1. Cada banda da imagem de satélite é convertida em uma matriz, formando as matrizes A e B de mesmo tamanho correspondentes às bandas 3 e 4 respectivamente. Cada posição nesta matriz representa um pixel da imagem de satélite e, por isso, o tamanho da matriz é proporcional à resolução da imagem de satélite.

Para cada valor a na posição x da matriz A e o valor b da mesma posição x na matriz B, o cálculo do NDVI é realizado através da Equação 1 com o valor de a substituindo IVP e V sendo substituído pelo valor de b. Como pode ser visto na Figura 1, uma matriz C de mesmo tamanho de A e B é gerada resultante do cálculo do índice NDVI nas matrizes A e B.

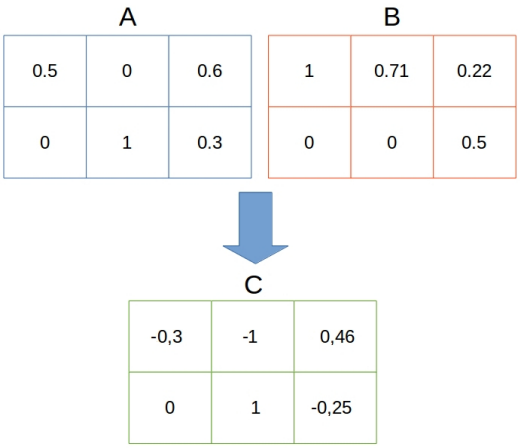


Figura 1. O cálculo do NDVI é feito em cada posição das matrizes A e B gerando a matriz C.

O cálculo do NDVI pode ser realizado de forma paralela, já que o cálculo em cada posição x das matrizes A e B pode ser feito de forma independente. No processamento paralelo utilizando a CPU, cada núcleo da CPU fica responsável por calcular o NDVI de um bloco correspondente nas matrizes A e B. O balanceamento destes blocos na CPU deve ser realizado de forma que nenhum núcleo da CPU seja o gargalo de processamento do algoritmo.

A GPU é uma ótima solução para processamento paralelo do NDVI devido a grande quantidade de núcleos de processamento presente. Os algoritmos na GPU de-

vem ser projetados de forma que tenhamos um número de *threads* da ordem de milhares. No cálculo do índice NDVI nas matrizes A e B, o algoritmo deve ser projetado de forma que cada *thread* seja responsável por processar uma posição ou um conjunto de posições das matrizes.

Embora o cálculo do NDVI seja inerentemente paralelo, várias questões devem ser levadas em consideração para que uma implementação paralela não sofra de baixo desempenho e falta de recursos, como, por exemplo, memória. Uma estratégia de particionamento e balanceamento dos dados é imprescindível para se obter bons resultados. Além disso, numa arquitetura *ManyCore* o acesso aglutinado e estando em sintonia com a hierarquia de memória é fundamental para um bom desempenho. A seguir, apresentamos nossa solução para essas questões.

4. PNDVI: Um novo algoritmo de processamento paralelo do índice NDVI para grandes imagens de satélite

Nesta Seção é apresentada a descrição da solução de processamento paralelo do cálculo do índice NDVI de grandes imagens de satélite. Esta solução, denominada PNDVI (Parallel NDVI), foi implementada de forma a utilizar todo o poder de processamento disponível lidando com os limites de memória da máquina. Se a máquina tiver alguma GPU disponível, esta será utilizada para calcular o NDVI devido ao seu alto poder de processamento e à grande quantidade de núcleos. Caso contrário, os núcleos da CPU serão utilizados para o cálculo.

O algoritmo PNDVI foi implementado utilizando a biblioteca GDAL para ler e escrever blocos de imagem do disco. O Algoritmo 1 apresenta a descrição do PNDVI. Ele inicia com a leitura de um bloco da imagem de entrada do disco para a memória da CPU. O tamanho destes blocos de imagem são definidos a partir da memória da CPU disponível na máquina, de forma que o bloco não ultrapasse o limite da memória.

Nas linhas 2 e 3 do algoritmo, as bandas 3 e 4 são lidas do bloco de imagem e armazenadas nas matrizes A e B respectivamente. Se a máquina tiver uma GPU disponível, o algoritmo de cálculo do NDVI é processado na GPU (linhas 5 a 12) e, caso contrário, é processado na CPU (linhas 15 a 18). O algoritmo de processamento paralelo na GPU é apresentado em detalhes na Subseção 4.2 e o algoritmo de processamento na CPU é apresentado na Subseção 4.1.

O processamento do NDVI gera um bloco resultante que é escrito no disco na posição correspondente do bloco de entrada. Desta forma, não é preciso armazenar toda a imagem de resposta na memória RAM para então gerar a imagem final de resposta. Se tiver mais blocos da imagem, o próximo bloco é lido do disco e o algoritmo, na linha 1, processa o novo bloco da imagem. Se não houver mais blocos da imagem para serem lidos, o algoritmo é finalizado.

4.1. Processamento paralelo do índice NDVI utilizando a arquitetura *MultiCore* (CPU)

O processamento do NDVI utilizando a arquitetura MultiCore consiste de utilizar dois ou mais núcleos da CPU para calcular o índice NDVI. O processamento paralelo utilizando a CPU é realizado com cada núcleo da CPU responsável por um bloco da imagem. No

algoritmo PNDVI na CPU, é criado um número de *threads* igual ao número de núcleos da CPU. Para que seja possível balancear o processamento nos núcleos da CPU, são gerados pedaços das matrizes A e B de entrada para balancear entre estas *threads*, onde cada *thread* fica responsável por um pedaço com tamanho igual ao do bloco de entrada.

Algoritmo 1: Algoritmo PNDVI

Entrada: $imagem$: referência para a imagem no disco
Saída: Imagem com o índice NDVI aplicado

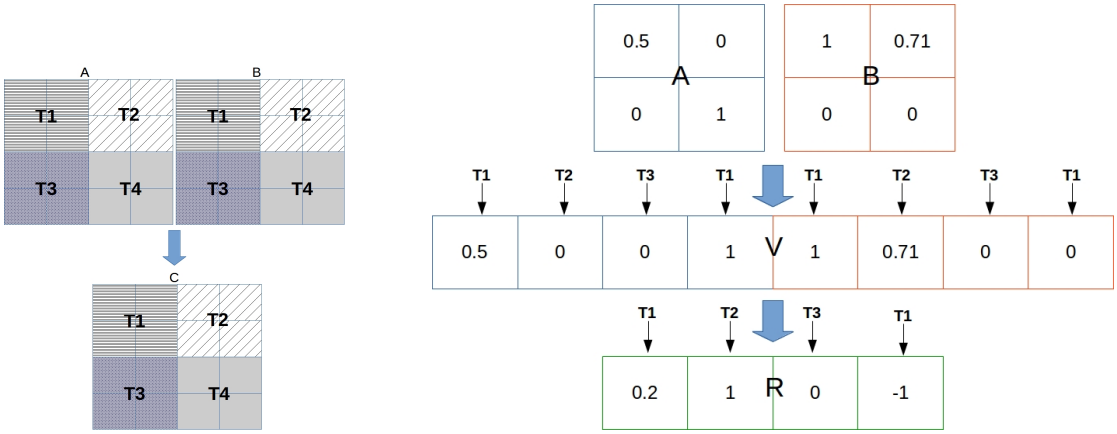
```
1  enquanto Tiver mais blocos da imagem no disco faça
2      A ← Banda 3 da imagem
3      B ← Banda 4 da imagem
4      se Existir uma GPU disponível na máquina então
5          V ← conteúdo das matrizes A e B
6          b ← tamanho da memória da GPU
7          enquanto Tiver mais blocos de tamanho b a partir de V faça
8              V' ← próximo bloco de tamanho b a partir de V
9              Transfere o vetor V' para a GPU
10             R ← Resultado do cálculo do índice NDVI na GPU
11             Transfere o vetor R da GPU para a CPU
12             Converte o vetor R na matriz C
13         fim
14     senão
15         B1 ← n blocos de tamanho b a partir da matriz A
16         B2 ← n blocos de tamanho b a partir da matriz B
17         Balanceia os blocos B1 e B2 entre as threads
18         C ← Cálculo o índice NDVI nos blocos B1 e B2 correspondentes
19     fim
20     Escreve a matriz C no disco
21 fim
22 retorna imagem de saída
```

Nas linhas 15 e 16 do Algoritmo 1 as matrizes A e B são divididas em n blocos iguais de tamanho b, onde n é o número de núcleos disponíveis na máquina. Na linha 17 estes blocos são balanceados entre as *threads*, de forma que cada *thread* fique responsável por pedaços iguais do bloco. Cada *thread* calcula o índice NDVI e armazena o resultado deste pedaço da imagem na matriz C (linha 18 do Algoritmo 1).

A Figura 2(a) apresenta um exemplo do algoritmo PNDVI na CPU. Cada *thread* fica responsável por processar blocos de mesmo tamanho das matrizes de entrada. A *thread* lê o pedaço das mesmas posições da matriz A e B e escreve o resultado na área correspondente da matriz C.

4.2. Processamento paralelo do índice NDVI utilizando a GPU

O algoritmo PNDVI foi implementado para ser processado na GPU utilizando a plataforma de computação paralela CUDA da NVIDIA. Com esta plataforma é possível paralelizar o processamento de algoritmos utilizando a grande quantidade de núcleos de processamento presentes na GPU. Na linha 5, o Algoritmo 1 recebe as bandas 3 e 4 car-



(a) Exemplo do algoritmo NDVI na CPU utilizando o algoritmo PNDVI

(b) Exemplo de processamento paralelo do NDVI na GPU utilizando o algoritmo PNDVI

Figura 2. Exemplo do algoritmo PNDVI na CPU e GPU.

regadas nas matrizes A e B respectivamente que são concatenadas em um único vetor V para que haja apenas uma transferência da CPU para a GPU.

Como a memória da CPU é geralmente maior que a memória da GPU, o vetor V pode não caber na memória da GPU. Por isso, são gerados blocos V' com tamanho b igual a dois terços da memória da GPU a partir do vetor V, já que um terço deve estar reservado para armazenar o vetor R resultante do processamento da GPU.

Entre as linhas 8 e 12 do Algoritmo 1, os blocos gerados são lidos dentro do vetor V. Na linha 12, o bloco V' é obtido a partir das primeiras b posições do vetor V. Os blocos seguintes são obtidos a partir das posições subsequentes. O vetor V' é transferido para a GPU, na linha 9, e o algoritmo NDVI executado na GPU (linha 10).

Dentro da GPU, o algoritmo inicia calculando o índice global da *thread* no processamento. Para aumentar o desempenho de acesso à memória global da GPU, cada *thread* acessa posições no vetor V' correspondentes ao índice global com *threads* acessando regiões próximas na memória global da GPU.

Como o tamanho do vetor V' de entrada pode ser maior que o número de *threads*, cada *thread* é responsável por calcular o índice NDVI de várias posições do vetor. Um *offset* é utilizado para que a *thread* possa obter posições do vetor que vão além do índice global da *thread*. Esse acesso aglutinado das *threads* otimizam o acesso à memória global da GPU, já que blocos podem ser lidos da memória global e utilizados por várias *threads* ao mesmo tempo.

Em cada iteração dentro da GPU, a *thread* obtém uma posição da primeira metade do vetor e a mesma posição na segunda metade do vetor. O resultado é calculado utilizando os dois valores obtidos e armazenado no vetor de resposta. No exemplo da Figura 2(b), o vetor V foi diretamente transferido para a GPU, pois o tamanho do vetor V é menor que a memória disponível na GPU. Como existem apenas três *threads* para executar o algoritmo a *thread* $T1$ fica responsável pelo cálculo de duas posições no resultado final.

A resposta é escrita no vetor R com o resultado do cálculo do NDVI. Este vetor R

é então transferido da GPU para a CPU na linha 11 do Algoritmo 1. Na linha 12, o vetor R é então convertido na matriz resultante C. Se houver mais blocos da imagem disponíveis no vetor V o algoritmo lê o próximo bloco e inicia novamente o algoritmo tendo este bloco como entrada. Se não houver mais blocos de entrada, a matriz C é retornada como resposta a CPU.

5. Avaliação de Desempenho

Diversos testes foram realizados para analisar o desempenho do algoritmo PNDVI na CPU e na GPU. Nos testes na CPU foi utilizada uma máquina com processador Intel Core i7-2670QM (2.2GHz, 6Mb Cache, 4 núcleos, 8 Threads) e 6GB de memória RAM. Na execução dos testes da GPU foi utilizada uma placa NVIDIA GeForce GT525M com 1GB de memória disponível e 96 CUDA *cores* com *clock* máximo de 1200 MHz.

Os testes na CPU foram realizados com até 8 *threads*. Como a complexidade de I/O da CPU e GPU é igual em relação à leitura e escrita do disco, o tempo de leitura e escrita dos blocos do disco foram desprezados do resultado final para analisar a diferença de desempenho entre a CPU e a GPU. O algoritmo implementado para ser processado na CPU foi desenvolvido na linguagem C++ e na GPU utilizando a linguagem CUDA da NVIDIA. Ambos algoritmos utilizaram a biblioteca GDAL do C++ para ler e escrever os blocos de imagem no disco.

Para cada imagem devem ser alocados o triplo do tamanho de cada banda da imagem. Dois terços deste espaço é utilizado para armazenar as bandas 3 e 4 da imagem e um terço para armazenar o resultado do processamento. Para representar cada pixel da imagem, a biblioteca GDAL aloca 8 bytes na memória RAM em uma estrutura denominada *GByte*. Por isso, cada imagem consome espaço na memória RAM igual a $r * c * 8$, onde r é o número de linhas da imagem e c o número de colunas.

Foram utilizadas quatro imagens de satélite com diferentes características nos testes: i) imagem A: cada banda com resolução de 3944 x 3041 e 95 MB de uso da memória RAM; ii) imagem B: cada banda com resolução de 7834 x 7059 e 442 MB de uso da memória RAM; iii) imagem C: cada banda com resolução de 8907 x 10454 e 744 MB de uso da memória RAM; iv) imagem D: cada banda com resolução de 25831 x 31208 e 6,5 GB de uso da memória RAM.

Para processar a imagem A devem ser alocados 285 MB de memória, e como as memórias da GPU e da CPU possuem tamanho maior que 285 MB, então a imagem A pode ser totalmente transferida para a GPU sem criar blocos menores. A Figura 3 contém o tempo de resposta do cálculo do índice NDVI na CPU e GPU. O processamento na CPU teve escalabilidade de aproximadamente 352% com o aumento de núcleos da CPU de 1 para 8. O tempo de resposta na GPU foi 2 vezes menor que o tempo de resposta da CPU com 8 núcleos e teve *speedup* de 7 vezes em relação ao tempo de resposta do algoritmo sequencial na CPU.

A Figura 4 apresenta o tempo de resposta do cálculo do NDVI das imagens B e C. Devem ser alocados 1,3 GB e 2,2 GB para processar as imagens B e C respectivamente. Como a GPU tem limite de memória de 1GB e a memória RAM limite de 6GB, estas imagens podem ser carregadas totalmente na memória RAM, mas devem ser dividas em blocos para serem processadas na GPU. Cada bloco de imagem tem tamanho aproximado

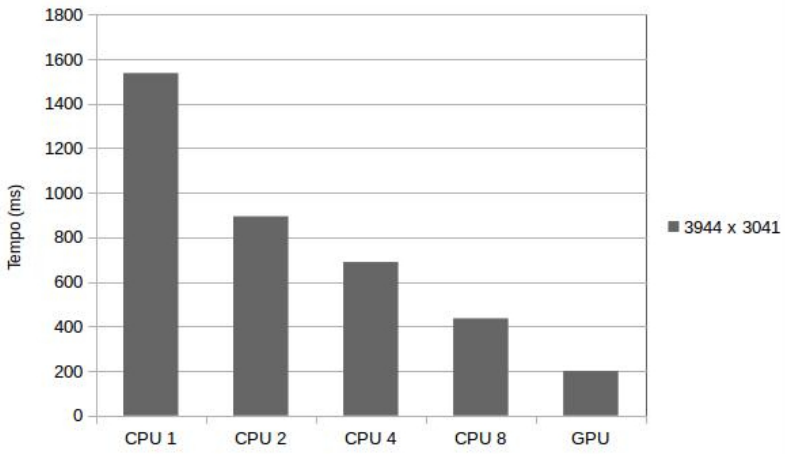
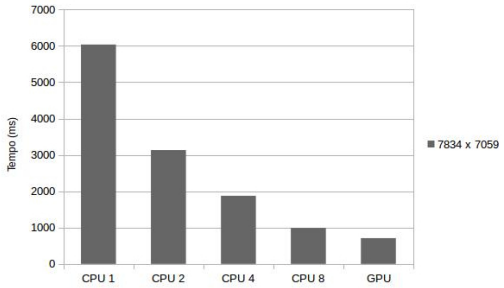


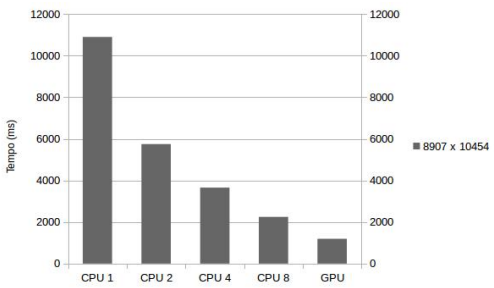
Figura 3. Gráfico com o resultado do cálculo do NDVI com a imagem A.

de 600 MB para armazenar as bandas 3 e 4. O espaço restante na GPU é utilizado para armazenar o resultado do processamento.

Como pode ser visto na Figura 4(a), o processamento da imagem B teve escalabilidade de aproximadamente 6 vezes com o aumento de *threads* de 1 para 8. O tempo de resposta na GPU foi 39% menor que o tempo de resposta na CPU com 8 *threads* e teve speedup de 8 vezes em relação ao algoritmo sequencial na CPU. O processamento da imagem C teve escalabilidade de aproximadamente 5 vezes com o aumento do número de *threads* de 1 para 8 (Figura 4(b)). A GPU teve desempenho 80% melhor que a CPU com 8 *threads* e *speedup* de 9 vezes em relação ao algoritmo sequencial da CPU.



(a) Resultado do cálculo do NDVI da imagem B



(b) Resultado do cálculo do NDVI da imagem C

Figura 4. Tempo de resposta do cálculo do NDVI de imagens que cabem na memória da CPU, mas não cabem na memória da GPU.

Para testar o caso onde a imagem não cabe na memória da CPU, foi utilizada a imagem D que possui resolução 25831 x 31208 com 6,5 GB de uso da RAM para cada banda da imagem e 19,5 GB de uso total da memória para armazenar as bandas 3 e 4, além do resultado final. Como a memória da CPU possui tamanho de 6 GB é necessário utilizar a abordagem de leitura de blocos de imagens do disco. Assim, são necessárias quatro leituras de blocos do disco, com cada bloco de 3,2 GB de tamanho. O restante da memória é utilizado para armazenar as estruturas intermediárias do algoritmo e o resultado final.

A GPU tem limite de 1 GB de memória e, por isso, foram gerados blocos de

tamanho de 600 MB divididos entre as bandas 3 e 4. O gráfico da Figura 5 apresenta o tempo de processamento em milissegundos do algoritmo PNDVI com a imagem D. Houve um ganho de aproximadamente 5 vezes com o aumento do número de *threads* de 1 para 8 na CPU. O tempo de processamento na GPU foi aproximadamente 53% menor que a CPU com 8 *threads*. A GPU teve um *speedup* de 9 vezes em relação ao algoritmo sequencial na CPU.

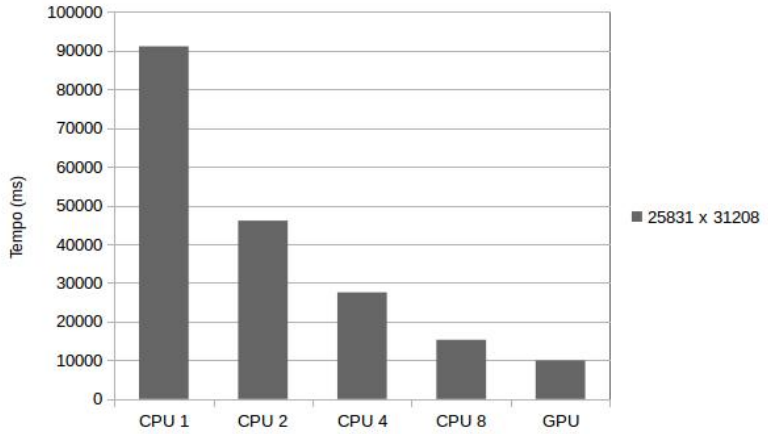


Figura 5. Gráfico com o resultado do cálculo do NDVI da imagem D que não cabe na memória RAM.

O *speedup* do algoritmo PNDVI foi de até 9 vezes no processamento de todas as imagens. Quanto maior a imagem, maior foi o ganho de *speedup* entre a CPU e a GPU, exceto o processamento da imagem D com maior resolução (25831 x 31208) que apresentou um *speedup* igual ao da imagem C. Isto ocorreu devido à grande quantidade de transferências de blocos da CPU para a GPU necessários para processar a imagem D. Portanto, quanto maior a memória disponível na GPU, melhor será o desempenho do algoritmo.

Os artigos apresentados na Seção 2 de trabalhos correlatos realizaram testes somente com imagens que cabem na memória da CPU deixando de tratar casos de imagem cujo tamanho seja maior que a memória disponível. Nenhum dos trabalhos correlatos apresentaram soluções de processamento paralelo na CPU e na GPU dependendo dos recursos disponíveis. No cálculo do NDVI, entre os trabalhos na revisão bibliográfica, o maior *speedup* foi obtido em [Juan and Jianchao 2013] cuja solução apresentou ganho de até 6 vezes entre a GPU e a solução sequencial.

Os testes apresentados neste trabalho foram construídos utilizando imagens com diferentes tamanhos, para demonstrar o desempenho do algoritmo paralelo PNDVI em cenários onde existe a necessidade de lidar com os limites de memória da CPU e da GPU. Houve um ganho significativo de desempenho do algoritmo PNDVI com o aumento do número de *threads* na CPU. A utilização da GPU obteve melhor desempenho que a CPU com 8 *threads* tendo até 9 vezes de *speedup* em relação ao algoritmo sequencial da CPU.

6. Conclusões

O índice NDVI é importante no estudo de vegetação, pois evidencia, a partir do uso de imagens de satélite, o vigor e a caracterização da vegetação de uma área. O cálculo

do NDVI é feito utilizando as bandas 3 e 4 das imagens a partir da diferença entre as reflectâncias destas bandas dividido pela soma das reflectâncias das duas bandas.

O volume de imagens de satélite cresceu bastante nos últimos anos e o tamanho de cada imagem gerada também tem crescido rapidamente com a melhor resolução destas imagens. Para processar de forma eficiente grandes imagens de satélite, o processamento deve ser realizado de forma paralela. As arquiteturas MultiCore, que utilizam mais de um núcleo da CPU, e a arquitetura ManyCore, com a grande quantidade de núcleos da GPU, têm sido amplamente utilizadas para processar de forma eficiente as imagens de satélite.

Para tratar grandes imagens de satélite devem ser implementadas estratégias que tratem imagens com tamanho maior que o limite de memória disponível.

Este trabalho apresentou uma solução eficiente para processamento paralelo do índice NDVI em grandes imagens de satélite na CPU e na GPU. A implementação proposta faz uso de uma técnica de particionamento e balanceamento de carga que permite que grandes imagens sejam processadas eficientemente de forma paralela em máquinas de baixo custo. Além disso, o uso do acesso aglutinado e hierarquias de memória na GPU resulta numa implementação com menor número de acesso à memória global da GPU e, consequentemente, menor complexidade de I/O.

O algoritmo proposto no trabalho, denominado PNDVI, permite processar grandes imagens através da leitura de blocos de imagens do disco que caibam na memória da CPU e da GPU. O algoritmo PNDVI apresentou uma redução do tempo na solução na CPU com o aumento do número de *threads* e *speedup* de até 9 vezes da solução na GPU em relação à solução centralizada com uma CPU.

Em trabalhos futuros, será implementada uma técnica de *streaming* de dados e uma versão multi-gpu para aproveitar os recursos computacionais presentes em um ambiente com várias GPUs. Além disso, serão conduzidos avaliações do tamanho do bloco e o consequente impacto da comunicação entre a CPU e GPU no desempenho do algoritmo PNDVI.

Referências

ACHHAB, N. B., RAISSOUNI, N., AZYAT, A., CHAHBOUN, A., and LAHRAOUA, M. (2010). Multithreading on multicores impact on remote sensing computations. *Mathematical Problems of Computer Science*, 2(12):7885–7895.

Alvarez-Cedillo, J., Herrera-Lozada, J., and Rivera-Zarate, I. (2014). Implementation strategy of ndvi algorithm with nvidia thrust. In *Image and Video Technology*, pages 184–193. Springer.

Chemin, Y. (2010). Gpgpu with gdal-basics of gpgpu interfacing. *OSGeo Journal*, 6(1):3.

Chen, D., Wang, L., Tian, M., Tian, J., Wang, S., Bian, C., and Li, X. (2013). Massively parallel modelling & simulation of large crowd with gpgpu. *The Journal of Supercomputing*, 63(3):675–690.

Christophe, E., Michel, J., and Inglada, J. (2011). Remote sensing processing: From multicore to gpu. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 4(3):643–652.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., and Turner, D. P. (2003). An improved strategy for regression of biophysical variables and landsat etm+ data. *Remote Sensing of Environment*, 84(4):561–571.

Gupta, A., Naidu, S. D., Srinivasan, T., and Krishna, B. G. (2011). A gpu based image matching approach for dem generation using stereo imagery. In *Engineering (NUiCONE), 2011 Nirma University International Conference on*, pages 1–5. IEEE.

Juan, W. and Jianchao, S. (2013). A new type of ndvi algorithm based on gpu dividing block technology. In *Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on*, pages 709–712. IEEE.

Liu, P., Yuan, T., Ma, Y., Wang, L., Liu, D., Yue, S., and Kołodziej, J. (2014). Parallel processing of massive remote sensing images in a gpu architecture. *COMPUTING AND INFORMATICS*, 33(1):197–217.

Liu, Y., Chen, B., Yu, H., Zhao, Y., Huang, Z., and Fang, Y. (2011). Applying gpu and posix thread technologies in massive remote sensing image data processing. In *Geoinformatics, 2011 19th International Conference on*, pages 1–6. IEEE.

Ma, Y., Zhao, L., and Liu, D. (2009). An asynchronous parallelized and scalable image resampling algorithm with parallel i/o. In *Computational Science–ICCS 2009*, pages 357–366. Springer.

Miura, T., Huete, A. R., Yoshioka, H., and Holben, B. N. (2001). An error and sensitivity analysis of atmospheric resistant vegetation indices derived from dark target-based atmospheric correction. *Remote sensing of Environment*, 78(3):284–298.

Nictua, I. and ALDEA, O. Using multiprocessor systems for multispectral data processing. *Pan*, 7:2–09.

Plaza, A. J. and Chang, C.-I. (2007). *High Performance Computing in Remote Sensing*. Chapman & Hall/CRC.

Plaza, A. J., Plaza, J., and Paz, A. (2010). Parallel heterogeneous cbir system for efficient hyperspectral image retrieval using spectral mixture analysis. *Concurrency and Computation: Practice and Experience*, 22(9):1138–1159.

Qu, H., Zhang, J., Lin, Z., and Chen, H. (2013). Parallel acceleration of sam algorithm and performance analysis. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(3):1172–1178.

Qu, X., Li, J., Zhao, W., Zhao, X., and Yan, C. (2010). Research on critical techniques of disaster-oriented remote sensing quick mapping. In *Multimedia Technology (ICMT), 2010 International Conference on*, pages 1–4. IEEE.

Saribekyan, A. G. (2013). A parallel algorithm for geoprocessing of vegetation indices. *Mathematical Problems of Computer Science*, 40(1):39–43.

Shrestha, B., O'Hara, C. G., and Younan, N. H. (2006). Strategies for alternate approaches for vegetation indices compositing using parallel temporal map algebra. In *Proceedings of the ASPRS 2006 Annual Conference, Reno, Nevada*, pages 1–11.

Thomas, U., Rosenbaum, D., Kurz, F., Suri, S., and Reinartz, P. (2009). A new software/hardware architecture for real time image processing of wide area airborne camera images. *Journal of Real-Time Image Processing*, 4(3):229–244.

Wang, Y., Ma, Y., Liu, P., Liu, D., and Xie, J. (2010). An optimized image mosaic algorithm with parallel io and dynamic grouped parallel strategy based on minimal spanning tree. In *Grid and Cooperative Computing (GCC), 2010 9th International Conference on*, pages 501–506. IEEE.

Zhu, H., Cao, Y., Zhou, Z., and Gong, M. (2012). Parallel multi-temporal remote sensing image change detection on gpu. In *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International*, pages 1898–1904. IEEE.

Segmentação e Identificação Automática de Placas de Aço

Rogério E. de Jesus, Warley R. Mendes, Karin S. Komati, Daniel S. Cavalieri

Programa de Pós-Graduação em Controle e Automação, PROPECAUT
Instituto Federal do Espirito Santo – Campus Serra – Serra, ES – Brazil

rejesus@outlook.com, warleymendes@gmail.com, kkomati@ifes.edu.br, daniel.cavalieri@ifes.edu.br

Abstract. *This paper presents a system's proposal for text localization and character segmentation in images for tracking slabs in the steel-making industry. The input data are images of slabs on a roller table in different lighting conditions. Since, there are several representations of color in digital images, each one with different characteristics concerned to illumination invariance, in this work two color space are compared: HSV and L*a*b*. The experimental results show that the proposed algorithms are reliable and segmentations based on L*a*b* color space present higher accuracy than the results of HSV color space.*

Resumo. *Este artigo apresenta uma proposta de um sistema de localização de texto e segmentação de caracteres em imagens para rastrear placas de aço na indústria siderúrgica. Os dados de entrada são imagens de placas em uma mesa de rolos em diferentes condições de iluminação. Uma vez que, existem várias representações de cores em imagens digitais, cada uma com diferentes características devido a invariância de iluminação, neste trabalho dois espações de cores são comparados: HSV e L*a*b*. Os resultados experimentais mostram que os algoritmos propostos são confiáveis e segmentações baseadas no espaço L*a*b* apresentam maior precisão do que os resultados do espaço de cores HSV.*

1. Introdução

O aço é uma liga de ferro e carbono, e passa por uma série de processos em uma indústria siderúrgica: Preparação da carga, Redução, Refino e Lingotamento e Laminação (Figura 1) [Instituto Aço Brasil 2015]. O aço é produzido conforme as características específicas para cada pedido de cada cliente, variando as dimensões e propriedades mecânicas, por exemplo.

Cada pedido entra em uma carteira de produção e o aço é fabricado para atender as especificações apresentadas. Após o lingotamento contínuo, o aço se torna uma placa e que recebe uma marcação alfanumérica que o identifica de maneira única - SlabID, a Figura 2 apresenta exemplos desta identificação. A placa de aço pode ser enviada para alguns clientes, ou ainda, pode seguir para outro processo chamado Laminador de Tiras à Quente (LTQ) e ficar na forma de bobina.

É importante salientar que a marcação do SlabID é executada por um operador humano, e muitas vezes o espaçamento entre os caracteres não é o mesmo, podendo ocorrer sobreposição dos caracteres, muitas vezes os caracteres podem estar desalinhados, e

dependendo da movimentação da placa, a identificação pode ficar de cabeça para baixo, como exemplificado na imagem da Figura 3. Outros fatores afetam a qualidade da leitura, tais como aa oxidações existentes na placa de aço, as diferentes tonalidades existentes na placa de aço, variações de luminosidades do ambiente de operação industrial de uma siderúrgica: indo desde saturações na imagem provocadas pelo sol até a baixa luminosidade.

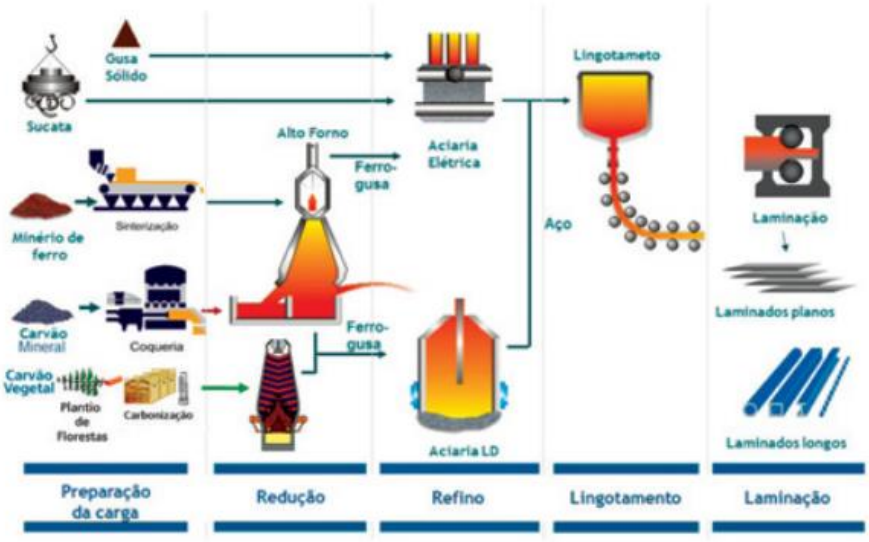


Figura 1. Fluxo simplificado de produção do aço de uma indústria siderúrgica

Figura 2. Fotos da identificação de placas na saída do lingotamento

Figura 3. Exemplo de imagem com difícil identificação da placa

Durante todo o processo de fabricação, é necessário manter o rastreamento do material para que, no final, o produto esteja em conformidade com as requisições do cliente, assim, inspeções visuais feitas pelos operadores humanos são utilizadas para o rastreamento das placas. Em função de fatores como: falhas de sensoriamento, falha de comunicação sistêmica ou fator humano, pode ocorrer uma falha no rastreamento de uma determinada placa, implicando em paradas do processo produtivo.

Uma alternativa para aumentar a robustez deste rastreamento é a inclusão do reconhecimento automático SlabID no processo de rastreamento das placas de aço, por

meio das imagens geradas por uma câmera fixa localizada no início do processo do LTQ. Neste sistema de monitoramento aplicam-se técnicas de processamento digital de imagens e visão computacional, localiza-se a região de texto da placa presente em cada quadro do vídeo e realiza-se o processo de classificação dos caracteres, rastreando a placa ao longo de seu deslocamento.

O presente trabalho baseia-se no trabalho de Jesus e Almonfrey, 2014, que realiza um estudo de localização de segmentação do ROI usando o espaço de cores HSV, obtendo índice de acerto de 75,5% numa base com 600 imagens. Pressupõe-se que um espaço de cores que seja mais insensível a variações de iluminação melhorará este índice de acerto. Selecionar o melhor espaço de cores ainda é uma das dificuldades na segmentação de cores [Severino, O. e Gonzaga 2013].

É importante ressaltar que, neste trabalho, o objetivo principal é realizar a localização e segmentação precisa de identificadores de placas de aço presentes em imagens sujeitas a diferentes condições de iluminação, capturadas em diferentes dias e em diferentes horários ao longo de um dia. A classificação dos caracteres propriamente dita, desde que se possua caracteres segmentados com qualidade, não é objetivo.

O texto encontra-se estruturado em quatro seções. Na Seção 2, são abordados os materiais e métodos empregados na solução do problema proposto neste trabalho. Na Seção 3, são discutidos os resultados obtidos com a implementação proposta e na, Seção 4, são apresentadas algumas conclusões da realização deste trabalho.

2. Materiais e Métodos

A abordagem utilizada neste trabalho inicia na Seção 2.1 com uma avaliação dos espaços de cores buscando características que permitam, segmentar a região de caracteres. Depois segue os mesmos passos descritos em Jesus e Almonfrey, 2014, descrevendo-se a etapa de pré-processamento com a utilização de alguns filtros preparando a imagem para a etapa posterior (na seção 2.2). Descreve a etapa de segmentação da região de texto na imagem (na seção 2.3), depois, a segmentação dos caracteres (na seção 2.4) e a classificação dos identificadores das placas de aço são realizadas (na seção 2.5). A solução do problema proposto neste trabalho foi desenvolvida utilizando-se o Matlab.

2.1. Avaliação do Espaço de Cores

Como a câmera empregada adquire imagens em um ambiente externo, características do estado da placa de aço, como carepa, ranhuras formadas pelo corte da placa no lingotamento contínuo influenciam na detecção da região de texto na imagem. Para tornar o processo de localização e detecção dos identificadores mais robusto quanto às influências mencionadas, foram avaliados dois espaços de cores: HSV e L*a*b*. No estudo anterior já se havia retirado o espaço de cores RGB por apresentar resultados piores que o HSV.

Diferentes espaços de cores representam uma imagem através de diferentes características [Cheng, Jiang, Sun e Wang, 2001]. Podemos encontrar através de um destes espaços, uma característica, ou um conjunto delas, que permita localizar a região de interesse (ROI) mais assertivamente. Assim, para diferentes imagens de placas, realiza-se um trabalho de exploração buscando encontrar características comuns a região

de caracteres em diferentes condições. A Figura 4 mostra a imagem de entrada colorida e as Figuras 5 e 6 apresentam a imagem de cada canal dos espaços de cores HSV e L*a*b*, respectivamente.

Figura 4. Imagem original

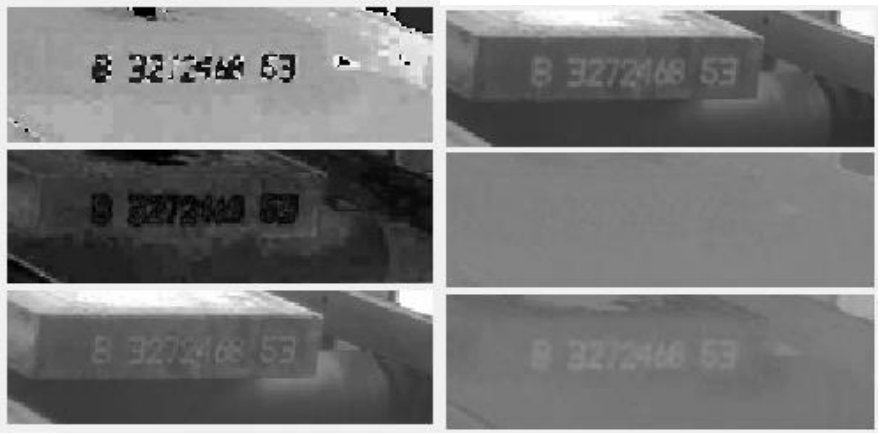


Figura 5. Canais HSV e L*a*b

O espaço de cores L*a*b* (CIELAB) é um espaço no qual a dimensão L representa a luminância e as dimensões a e b para as cores. Um dos mais importantes atributos deste espaço de cores é que é independente de dispositivo. A Figura 7 mostra o espaço de cores L*a*b*, e observa-se que para a cor preta, L* = 0, e para o branco L* = 100. Quando os canais de cores são a* = 0 e b* = 0, então as cores são cinza. As cores vermelho e verde são opostas ao longo do eixo a*, sendo mais verde para valores negativos de a* e mais vermelho para valores positivos de a*. Para o eixo b*, quanto mais negativo, mais azul e quanto mais positivo, mais amarelo é a cor representada. Os limites das dimensões de a* e b*, em geral, variam de ±100.

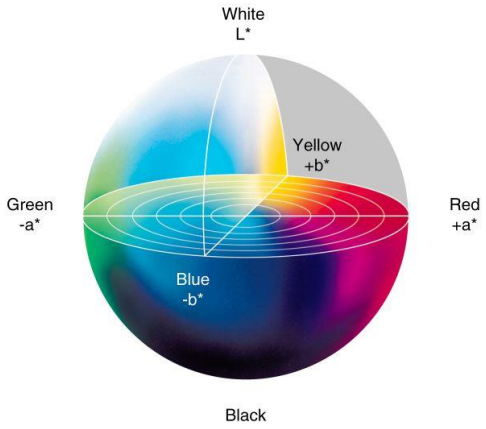


Figura 7. Espaço de Cores L*a*b*.

Outro resultado preliminar é que os parâmetros de filtragem para cada canal nos diferentes espaços de cores devem ser feitos de forma diferenciada para imagens diurnas e noturnas. Verifica-se que as imagens diurnas possuem maior luminosidade (L*) como mostrado na Figura 8, permitindo assim identificar e utilizar diferentes ajustes para os filtros. Para classificar uma imagem como noite, todos os valores do canal L eram menores que 65. Este valor foi obtido empiricamente, por inspeção em um conjunto reduzido de imagens aleatórias da base de dados de testes.

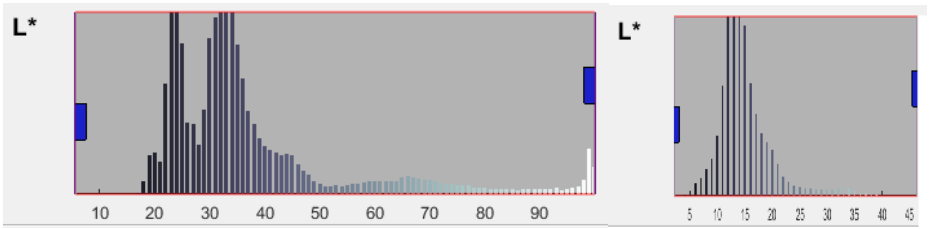


Figura 8. Luminosidade (L*) para imagens diurnas e noturnas respectivamente.

2.2. Pré-processamento da Imagem

Nesta etapa prepara-se a imagem visando eliminar características que não interessam a cena, tais como oxidações existentes na placa de aço, diferenças de tonalidades existentes na placa de aço, pouco contraste na imagem, marcação dos caracteres com variação de cor ou forma, variações de luminosidades no ambiente de operação, saturações na imagem provocadas pelo sol e baixa luminosidade, tudo isto afeta a obtenção da ROI.

Considerando-se a imagem de entrada como a Figura 4, e utilizando o espaço de cores L*a*b*, aplicamos uma técnica de binarização com limiares diferenciados para imagens diurnas e noturnas como indicado nas Figuras 9 e 10 (importante observar que os limites superiores do eixo x são diferentes nestes gráficos).

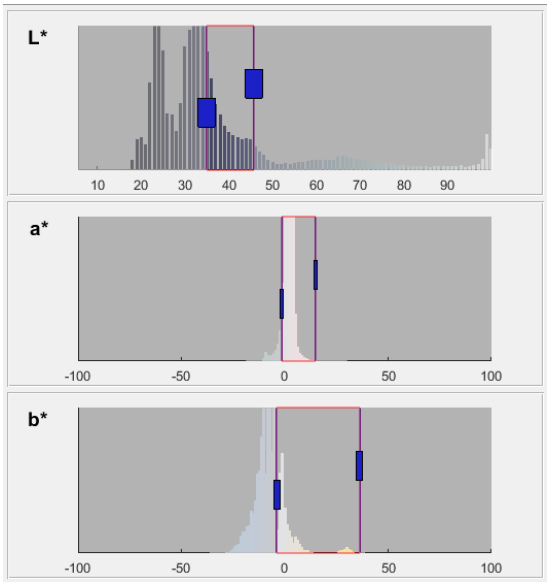


Figura 9. Filtragem utilizada para imagens diurnas.

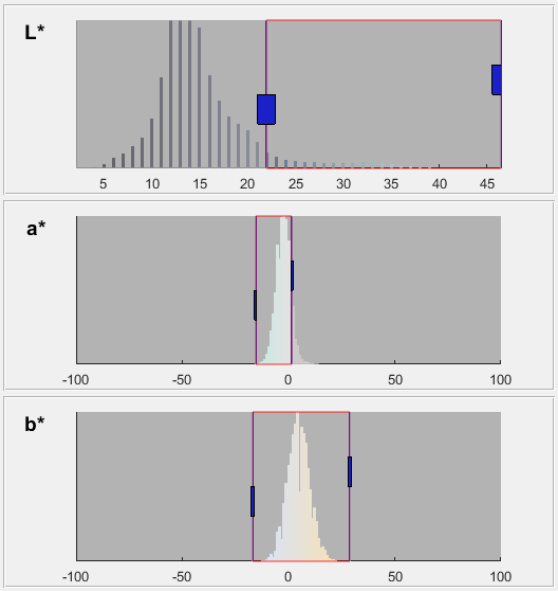


Figura 10. Filtragem utilizada para imagens noturnas.

As janelas (retângulos) demarcados, representam o filtro aplicado a cada dimensão do espaço de cor L*a*b*. Tudo que estiver dentro do retângulo é branco e o que estiver fora dos retângulos se torna preto. Na figura 9, a filtragem usada para imagens em dia. Na 10, em imagens noturnas. O algoritmo identifica qual delas usar pela escala da luminosidade obtida (figura 8). Os valores foram obtidos empiricamente, por inspeção em um conjunto reduzido de imagens aleatórias da base de dados de testes. Forma similar ao HSV.

Na Figura 11, temos o resultado obtido pela aplicação do filtro. Percebe-se que a maioria das informações que não interessam foram removidas da imagem.

Figura 11. Imagem pós-processada pela filtragem.

Contudo, a imagem obtida ainda apresenta ruídos que devem ser removidos. Aplica-se, então, o filtro da mediana [Gonzales e Woods, 2000], uma transformação bastante conhecida para suavizar ruídos do tipo impulsivo, ou sal e pimenta, em sinais e imagens digitais. A imagem binarizada e suavizada obtida com a aplicação do filtro da mediana se encontra na Figura 12.

Figura 12. Imagem após filtr.o da mediana

Em seguida, aplica-se um filtro Sobel para extração de bordas, obtendo-se imagem apresentada na Figura 13, com o contorno dos objetos.

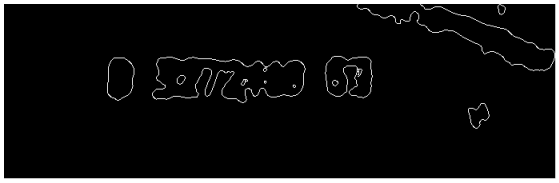


Figura 13. Imagem pós-filtro de Sobel.

2.3. Segmentação da Região de Texto na Imagem

A etapa de segmentação da região de texto tem por objetivo fazer um recorte na imagem para a extração da ROI da imagem. Para esta etapa utiliza-se o mesmo algoritmo apresentado por Jesus e Almonfrey, 2014.

A partir da imagem com bordas obtida na Seção B, deve-se adotar uma estratégia para identificar a região onde caracteres se encontram na imagem. É possível perceber que a ocorrência de bordas tende a ser mais intensa na região onde caracteres se encontram. Pode-se medir esta característica utilizando, por exemplo, a soma dos pixeis ao longo de um conjunto de linhas da imagem. Com este objetivo, para cada linha da imagem à esquerda da Figura 14, executa-se este cálculo, obtendo-se um gráfico conforme apresentado à direita na mesma figura. Este gráfico será chamado ao longo do texto de soma vertical, obtido a partir de cada linha da imagem.

Com o perfil da soma da Figura 14, é possível se obter os limites verticais da região onde o texto se encontra na imagem, ou seja, as linhas superior e inferior da matriz imagem. Calcula-se o valor médio das somas ajustado por um fator multiplicativo empírico, 0.9 no caso, e varre-se cada posição de altura do gráfico, correspondente a soma dos pixeis de cada linha da imagem, até que seja encontrado um valor maior que a média em uma transição positiva. Ao mesmo tempo, verifica-se a ocorrência de uma transição negativa da linha (altura) i para a linha i+1. Caso a distância entre a posição da transição positiva e a posição da transição negativa esteja dentro de limites estabelecidos, considera-se a região encontrada como candidata válida. Caso estes limites estejam dentro de um intervalo de distância, em relação à mediana encontrada, considera-se que a região de interesse, os limites superior e inferior, foram encontrados.

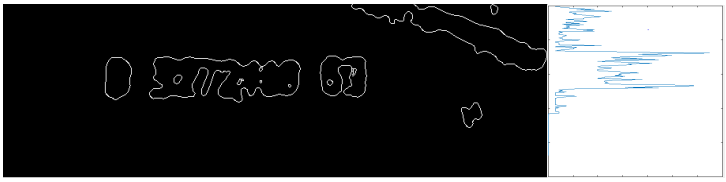


Figura 14. Imagem com bordas e o gráfico da soma vertical (altura [pixel] x soma [pixeis]).

Na Figura 15, exibem-se os valores de média (linha vermelha horizontal), limites (linhas vermelhas verticais) e da mediana (linha verde) obtidos para o gráfico da Figura 14, só que agora apresentado na horizontal.

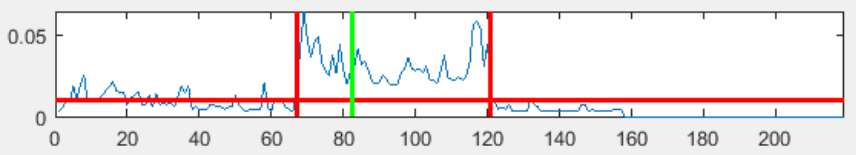


Figura 15. Gráfico da soma vertical com limites identificados – (soma [pixel] x altura [pixel]).

A partir dos limites verticais obtidos, pode-se selecionar a área de interesse vertical recortando, na imagem, à região que contém os caracteres (Figura 16).

Figura 16. Imagem obtida após a seleção da região de interesse na vertical.

Analogamente ao processo realizado para cada linha da imagem, para cada coluna da imagem calcula-se a soma dos pixeis obtendo-se o gráfico indicado na Figura 17. Chama-se este gráfico, ao longo do texto, de soma horizontal obtido a partir de cada coluna da imagem.

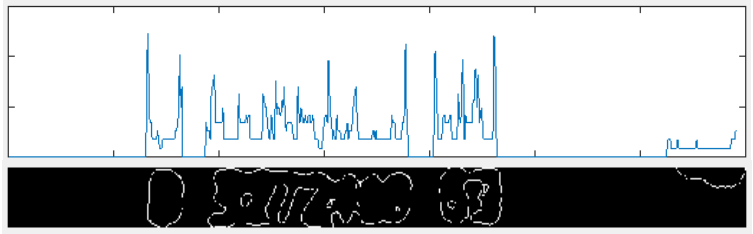


Figura 17. Imagem com bordas e o gráfico da soma horizontal (soma [pixeis] x largura [pixel]).

Um algoritmo semelhante ao aplicado ao vetor de somas vertical é utilizado. Mas, neste caso, algumas adaptações são necessárias em função da existência de um menor número de bordas por coluna da matriz imagem, tornando a análise de somas nas colunas da imagem mais suscetível a ruídos. A varredura é realizada nos dois sentidos. Três diferentes valores são encontrados considerando a comparação com a mediana, sendo mantido o menor valor no caso do limite inferior e o maior valor no caso do limite superior. O resultado é apresentado na Figura 18. A partir dos limites obtidos, pode-se restringir a área de interesse horizontal, caso ela tenha dimensões esperadas para os identificadores das placas de aço, completando a localização, na imagem, da região que contém os caracteres (Figura 19). Com esta região, é possível agora se realizar a segmentação e posterior classificação dos caracteres.

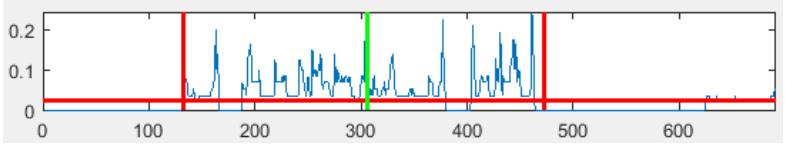


Figura 18. Gráfico de soma horizontal com limites identificados (soma [pixeis] x largura [pixel]).

Figura 19. Região de interesse identificada pelo algoritmo.

2.4. Segmentação dos caracteres

Nesta etapa, será realizada a segmentação e a classificação dos números presentes nos identificadores das placas de aço. A partir da ROI encontrada, separa-se esta região (Figura 20). Deseja-se que os caracteres alfanuméricos sejam separados do restante da imagem, sendo esta etapa denominada de segmentação dos caracteres.

Figura 20. Imagem originada da ROI (Region of Interest).

Observando-se o histograma da imagem transformada para tons de cinza, apresentado na Figura 21, observa-se uma característica bimodal. Neste caso, o método de binarização de Otsu pode ser utilizado com bom resultado. Esta técnica encontra um limiar ótimo, aproximando o histograma por duas funções gaussianas, de forma a minimizar a variância intra-classes. A linha vertical vermelha nesta figura representa o limiar encontrado pelo método de binarização.

A partir do limiar retornado pelo método de Otsu [7], pode-se realizar a binarização da imagem em tons de cinza, obtendo-se, assim, a imagem apresentada na Figura 22.

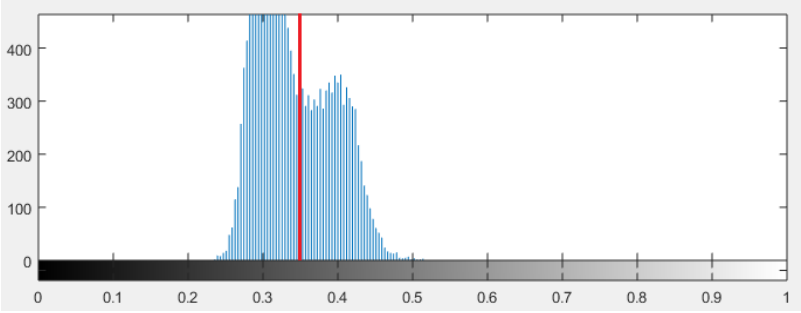


Figura 21. Histograma da imagem ROI em escalas de cinza (no de ocorrências x tons de cinza).

Figura 22. Imagem binarizada separando os caracteres.

2.5. Reconhecimento dos caracteres

Com a imagem binarizada obtida, pode-se utilizar um OCR (Optical Character Recognition), que seja capaz de reconhecer caracteres a partir de um arquivo de imagem. Com isto, o objetivo final do trabalho será alcançado, pois, a partir da imagem do identificador alfanumérico presente na placa de aço, pode-se obter um arquivo de texto com as informações dos caracteres presentes nesse identificador.

A solução OCR utilizada, neste trabalho, é o próprio classificador do Matlab que utiliza como engine uma aplicação disponível de forma livre pelo Google chamada Tesseract [Smith 2007]. Apresenta-se ao classificador uma imagem como a da Figura 22 e obtêm-se o resultado da classificação da sequência de caracteres reconhecidos.

Para a imagem da Figura 22 o resultado obtido foi: “8 32724&8 53”. Pode-se observar que o resultado apresentou dois erros, o primeiro caractere foi incorretamente classificado como 8 sendo B, e o sétimo caractere foi incorretamente classificado como & sendo 6.

3. Resultados

A mesma base utilizada em Jesus e Almonfrey foi utilizada aqui de forma a poder comparar os resultados. Com 600 imagens em diferentes condições de horário e clima. Deste total, considerou-se 230 imagens em condições muito desfavoráveis: incidência de sol em apenas metade dos números dos identificadores da placa de aço, marcação não nítida ou manual, marcação com cores diferentes.

Na Tabela 1 apresenta-se o resultado obtido onde a coluna HSV representa o trabalho anterior utilizando este espaço de cores e L*a*b* representa o presente trabalho. Percebe-se um aumento no índice de acerto em qualquer condição. Em imagens boas o índice se eleva levemente de 93,80% para 95,68%, mas o aumento foi mais significativo para imagens consideradas ruins onde o valor se eleva de 46,10% para 64,35%. No total, o acerto se eleva de 75,5% para 83,67%.

Tabela 1. Índice de acerto da ROI

		HSV		L*a*b*	
Tipo da imagem	Número de imagens	ROI correta	Índice de acerto	ROI correta	Índice de acerto
boa	370	347	93,8%	354	95,68%
ruim	230	106	46,1%	148	64,35%
total	600	453	75,5%	502	83,67%

Na Figura 23, alguns bons resultados são apresentados, pode-se notar que o algoritmo foi eficiente para identificar a região de interesse em seis diferentes casos de iluminação e marcação da placa, que antes em HSV erravam e com o L*a*b* ficam corretos. No entanto, ainda há casos como os mostrados na Figura 24, no qual em três casos desfavoráveis a região foi identificada de forma incorreta.

Figura 23. Exemplos de obtenção da região de interesse pelo algoritmo.

Figura 24. Exemplos de obtenção da região de interesse pelo algoritmo de forma incorreta.

A acurácia do classificador não foi analisada por não ser o foco deste trabalho, pois utillizou-se um engine já existente. Entretanto, a técnica da obtenção do caractere mais recorrente dentre todos os quadros de um vídeo proposta por Jesus e Almonfrey, elegendo um vencedor para cada posição, já demonstrou ser eficiente na eliminação de falhas do classificador.

4. Conclusão

A aplicação de identificação textual em imagens ainda é desafiadora em situações específicas e em ambientes pouco estruturados, mas apresenta resultados que trazem benefícios ao processo produtivo industrial, como no caso abordado neste trabalho. Identificar o número da placa de aço presente na entrada do processo do LTQ pode evitar paradas indevidas no processo e auxilia o operador no processo de conferência dos identificadores das placas reduzindo seu esforço.

Neste trabalho, faz-se um estudo entre dois espaços de cores: HSV e L*a*b* sobre uma proposta já existente. Os resultados mostraram que o espaço de cores L*a*b* e melhor que o espaço HSV. Além disso, é feita uma diferenciação de ajustes para diferentes

períodos do dia e noite, e com isso alcançando um resultado eficiente na obtenção da região de interesse onde estão os caracteres na imagem.

Um índice de 95,68% de acerto foi alcançado em imagens sem interferências severas e um percentual de acerto de 64,35% foi alcançado em imagens desfavoráveis havendo ainda espaço para evolução, tais como a troca dos filtros utilizados por filtros adaptativos, futuramente usar um OCR mais moderno com maior acurácia e incluir estratágia baseada em HOG ou SIFT para localização do texto e comparação com outros sistemas [Sumathi, Santhanam e Devi, 2012].

Como trabalhos futuros, pode-se avaliar a utilização de outras técnicas, como redes neurais, para o reconhecimento dos caracteres e ampliar os estudos de trabalhos correlatos e outras metodologias encontradas na literatura.

Referências

Instituto Aço Brasil (Org.). Processo Siderúrgico: Etapas. 2015. Disponível em: <http://www.acobrasil.org.br/site2015/processo.html>. Acesso em: 01 maio 2016.

Jesus, R. e Almonfrey, D. (2014) “Segmentação e classificação de caracteres em placas de aço”. In Proc. of XX Congresso Brasileiro de Automática, pp. 1254-1257.

Severino, O. e Gonzaga, A. (2013) "A new approach for color image segmentation based on color mixture". Machine Vision and Applications. Vol. 24, nº 3, pp. 607-618.

Cheng, H. D., Jiang, X. H., Sun, Y. e Wang, J. (2001) "Color Image Segmentation: Advances and Prospects", Pattern Recognition, Vol. 34, nº 12, pp. 2259-2281.

Wang, X., Hänsch, R., Ma, L. e Hellwich, O. (2014) "Comparison of different color spaces for image segmentation using graph-cut," In Proc. of International Conference on Computer Vision Theory and Applications (VISAPP 2014), Lisbon, Portugal, pp. 301-308.

Gonzales, R. C. e Woods, R. E. (2000) “Processamento Digital de Imagens”, São Paulo: Edgard Blücher. p.137-138.

Otsu, N. (1979) "A Threshold Selection Method from Gray-Level Histograms," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66.

Smith, R. (2007) “An Overview of the Tesseract OCR Engine”. In: Proc. of the Ninth Int. Conference on Document Analysis and Recognition (ICDAR 2007). pp. 629–633.

Choi, S., Yun, J.P., Koo, K. e Kim, S. W. (2012) “Localizing slab identification numbers in factory scene images”. Expert Systems with Applications, Vol. 39, pp. 7621-7636.

Jung, K.; Kim, K. e Jain, A. (2004) “Text information extraction in images and video: A survey”. Pattern Recognition Letters. 13(1), p.977-997.

Sumathi, C. P., Santhanam, T. e Devi, G. G., (2012) “A survey on various approaches of text extraction in images”. International Journal of Computer Science and Engineering Survey 3 (4), 27

Um Estudo de Caso para a Segmentação de Falanges em Radiografias Carpais

Ding Yih An[1], Warley Rocha Mendes[1], Sérgio Nery Simões[2], Luiz Alberto Pinto[1], Karin Satie Komati[1]

[1]Programa de Pós-Graduação em Controle e Automação, PROPECAUT
Instituto Federal do Espírito Santo – Campus Esrra – Serra, ES – Brasil

[2]Coordenadoria de Informática
Instituto Federal do Espírito Santo – Campus Esrra – Serra, ES – Brasil

{sidnha,warleymendes,sergionery}@gmail.com,
{luiz.pt,kkomati}@ifes.edu.br

Abstract. *This article presents a case study on phalanges segmentation in hand-wrist x-ray radiographs images. In the used method, firstly, the radiographs images of the hands were segmented in five fingers sub-images. After that, the side profiles of the bone structure of each finger sub-images, originated in the first step, were mapped, and the proximal, middle and distal phalanges were segmented. Performance analysis used the 58 scanned x-ray radiographs images of the Greulich & Pyle atlas, and considered the evaluation of Precision, Accuracy and Error metrics. Average values, respectively, 97.83%, 98.20% and 1.80%, show that the proposed method can be a viable alternative to the phalanges segmentation problem.*

Resumo. *Este artigo apresenta um estudo de caso de segmentação de falanges em radiografias carpais. No método utilizado, inicialmente as imagens radiográficas das mãos foram segmentadas em cinco subimagens dos dedos. Em seguida, a partir das subimagens da fase anterior, os perfis laterais das estruturas ósseas dos dedos foram mapeados, e as falanges proximais, médias e distais foram segmentadas. A avaliação do desempenho utilizou as 58 imagens de Raios-x digitalizadas do atlas de Greulich & Pyle, e se baseou na análise das métricas Precisão, Exatidão e Erro. Os valores médios obtidos, respectivamente, 97,83%, 98,20% e 1,80%, mostram que o método proposto pode ser uma alternativa viável para a segmentação de falanges.*

1. Introdução

Técnicas de Processamento Digital de Imagens têm sido aplicadas na medicina para auxiliar na elaboração de diagnósticos, tais como em Pádua (2008), Gonçalves (2010) e Macedo (2012). Em particular, a segmentação de imagens [Gonzalez 2010] tem sido útil na clínica médica por oncologistas para a delimitação de tumores como em Souto (2014) e em Soares (2008), bem como por ortopedistas na delimitação de falanges em radiografias carpais para a realização de exames de estimação da idade óssea, Hsieh (2011) e Hsieh (2012). Nesse contexto, o exame para a estimação da idade óssea visa verificar o grau de desenvolvimento de um indivíduo, bem como diagnosticar doenças, distúrbios e anomalias do crescimento.

A segmentação de falanges consiste na delimitação dos contornos das estruturas ósseas dos dedos (falanges proximais, médias e distais, bem como as epífises) em

imagens de radiografias carpais, como a mostrada na Figura 1. Tal tarefa tem se mostrado desafiadora, entre outros motivos, devido à distribuição irregular dos raios-X na imagem radiografia, conhecido como Efeito Heel [Behiels 2002]. Como consequência, algumas regiões da imagem são pouco sensibilizadas, enquanto outras são sensibilizadas em excesso, comprometendo o contraste para a obtenção de imagens de qualidade [Olivete 2005]. Além disso, em muitos casos, as imagens apresentam apenas diferenças sutis entre as estruturas ósseas que se desejam segmentar e o tecido mole, Figura 1. Assim, o resultado final da segmentação das falanges pode apresentar erros consideráveis, em relação à região de interesse.

Figura 1. Radiografia carpal

Nesse contexto, o presente trabalho, ainda em fase inicial, apresenta um estudo de caso sobre segmentação de falanges em imagens radiográficas carpais e está constituído pelas seguintes seções: a Seção 2, Metodologia e Desenvolvimento, descreve as duas fases da abordagem utilizadas para a segmentação. Deve ser destacado que o método para segmentação das falanges e epífises descrito na segunda fase é uma contribuição original dos autores. A Seção 3, Resultados, apresenta os resultados das duas fases da metodologia aplicadas à segmentação das imagens do atlas de Greulich & Pyle (1992). A Seção 4, Considerações Finais, descreve as conclusões relativas ao estudo. A Seção 5, Trabalhos Futuros, relaciona as potenciais pesquisas que poderão ser futuramente desenvolvidas com temas correlatos ao aqui apresentado. Por fim, a Seção Bibliografia relaciona os trabalhos que embasaram o desenvolvimento do estudo.

2. Metodologia e Desenvolvimento

O trabalho foi desenvolvido em duas fases que, em conjunto visam à segmentação da imagem da estrutura óssea da região da mão, a partir de Radiografias Carpais. Na primeira fase, a imagem original, Figura 1, foi segmentada em cinco subimagens individuais, cada uma com um dos dedos. Na segunda, tendo como entrada as subimagens obtidas na fase anterior, os contornos das estruturas ósseas das falanges proximais, medias e distais foram mapeados e segmentados. A Figura 2 apresenta a arquitetura do sistema.

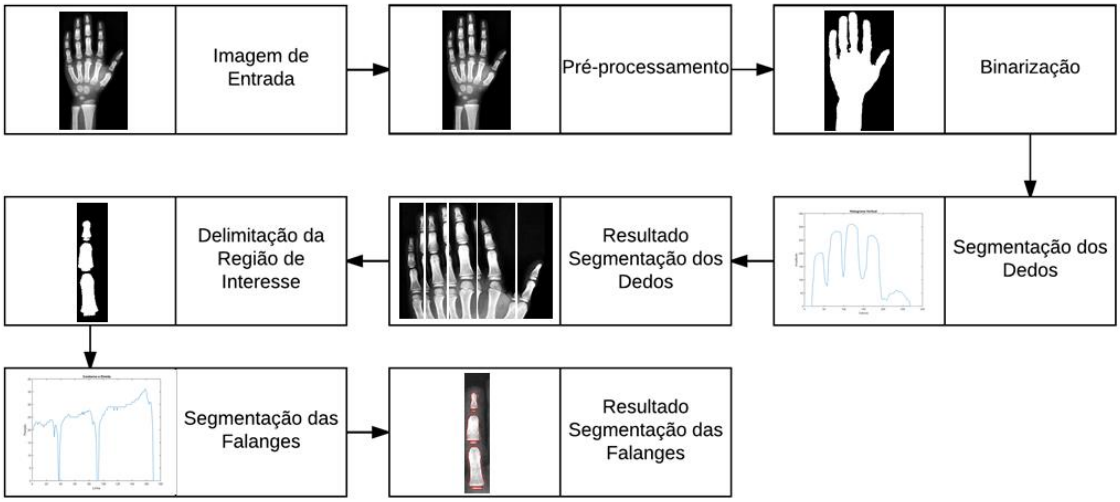


Figura 2. Arquitetura do Sistema

2.1. Primeira Fase – Segmentação dos Dedos

Nessa fase do trabalho, foram utilizadas as 58 imagens digitalizadas do atlas de Greulich & Pyle, disponibilizadas online em forma de slides pelo Dr. Geraldo Santana (2010), sendo 31 radiografias de mãos de indivíduos do sexo masculino, na faixa etária compreendida entre recém-nascido até 19 anos de idade; e 27 de indivíduos do sexo feminino, de recém-nascido até 18 anos de idade.

Pré-Processamento: Inicialmente, as imagens passaram por uma etapa de melhoria do contraste para a correção das não uniformidades inseridas pelo Efeito Heel [6]. Importante destacar que sem a correção do contraste, o desenvolvimento de métodos de segmentação baseados em Processamento Digital de Imagens, capazes de distinguir os ossos dos tecidos e do fundo [Castro 2009] pode se tornar difícil. O ajuste do contraste foi executado com a aplicação da função *Imadjust* do Matlab®. A Figura 3(a) mostra uma imagem original do atlas de Greulich & Pyle, e 3(b) mostra a mesma imagem realçada. Pode ser notada a significativa melhoria no contraste, e o realce das regiões da estrutura óssea em relação aos tecidos e ao fundo, o que pode melhorar o desempenho de algoritmos de binarização.

Binarização: Após o ajuste do contraste, as imagens foram convertidas do modelo RGB para HSV. Uma vez que o canal V no espaço HSV está associado à intensidade dos pixels, esse componente foi isolado e binarizado para efeito de segmentação. Pode-se argumentar que essa opção possibilita uma melhor discriminação da região de interesse em relação ao fundo. A binarização consistiu na aplicação da função *im2bw* do Matlab®, com um valor de limiar de 0,2 definido empiricamente. A

Figura 3(c) apresenta o resultado da operação, onde pode ser notada a boa definição do contorno da mão.

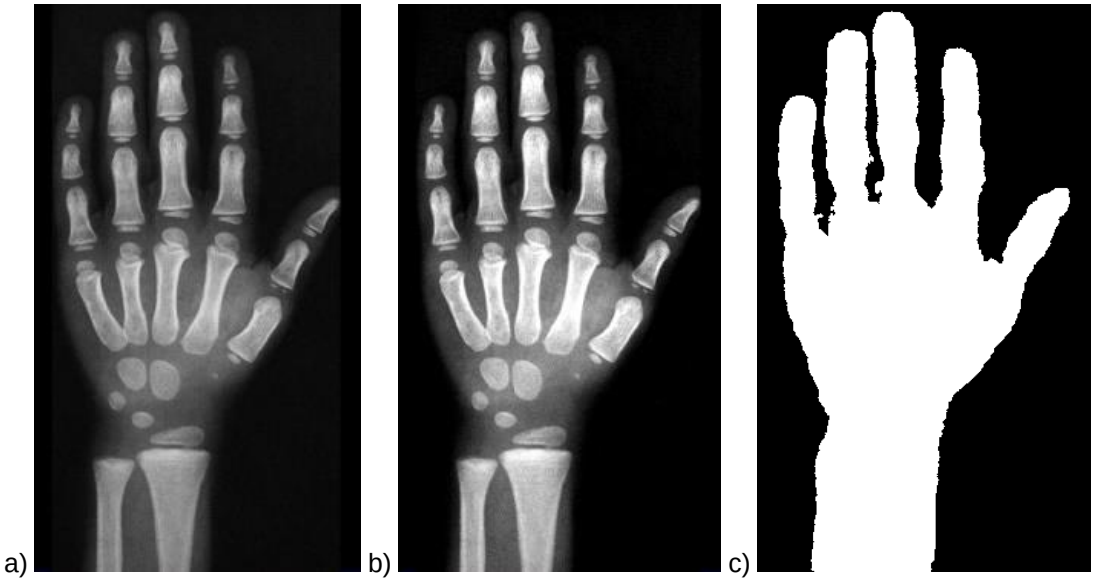


Figura 3. (a) Imagem de Radiografia (b) Imagem Ajustada (c) Imagem Binarizada

Segmentação: Para a segmentação das imagens binarizadas nas cinco subimagens dos dedos, foi aplicado o método utilizado em Castro (2009). Inicialmente foi traçado o perfil da mão com base na distribuição de pixels por coluna, aqui denominado histograma vertical. A Figura 4 mostra o gráfico que representa o histograma vertical da imagem da Figura 3(c). O eixo horizontal (*Coluna*) representa a posição da coluna na imagem, e o eixo vertical (*Amplitude*) indica o número de pixels com valor "1" na coluna correspondente.

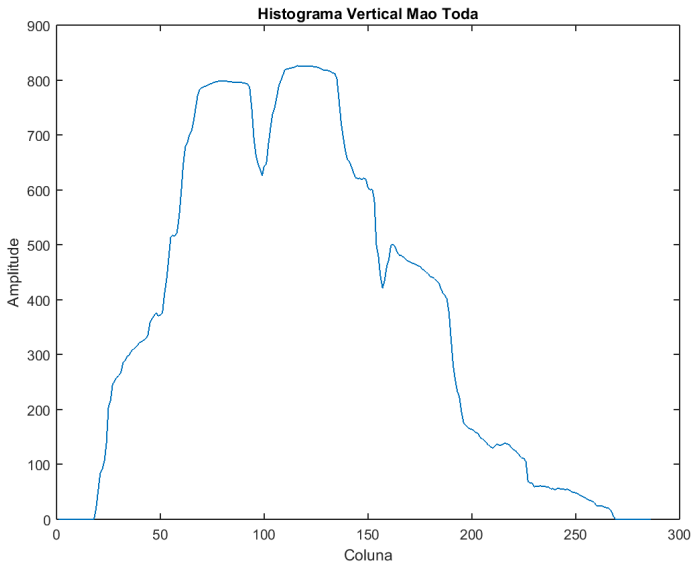


Figura 4. Histograma Vertical da Mão Binarizada

Observando o formato do histograma vertical na Figura 4, nota-se que o mesmo não reproduz o contorno da mão da Figura 3(c). Isso se deve a influência das regiões do punho e palma da mão, não diretamente associadas à região de interesse (dedos). Para

resolver esse problema, Castro propôs que as regiões do punho e palma fossem removidas por um corte horizontal. A definição da linha de corte baseou-se na análise da distribuição de pixels por linha, denominado histograma horizontal, e foi definida como a linha correspondente à maior amplitude, isto é, a linha com maior quantidade de pixels "1". Na Figura 5, o eixo horizontal (*Linha*) representa as linhas da imagem, e o eixo vertical (*Amplitude*) representa a quantidade de pixels "1" da linha correspondente.

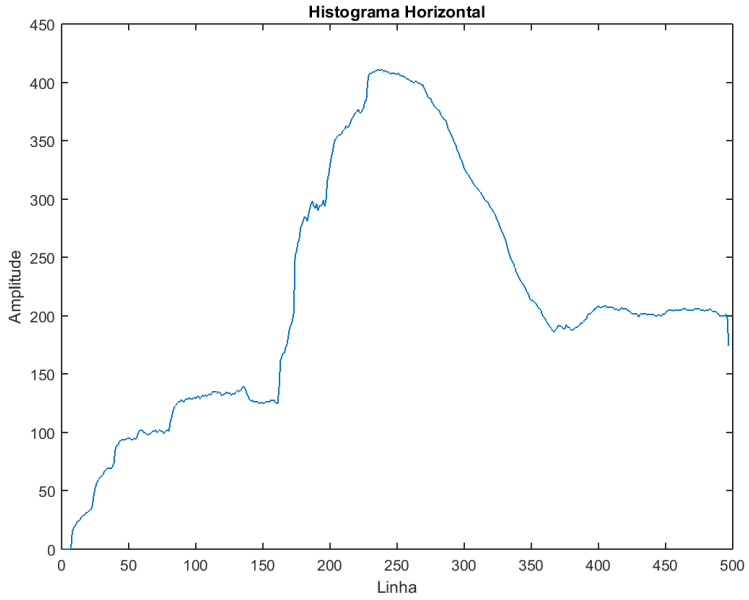


Figura 5. Histograma Horizontal da Mão Binarizada

Para a definição da linha de corte foi considerado um intervalo de segurança de 50 linhas abaixo da que representa o pico no histograma horizontal. A Figura 6 mostra o resultado do corte, e a eliminação das regiões sem interesse para a segmentação.

Figura 6. Imagem Cortada

Um novo histograma vertical, que está mostrado na Figura 7, foi obtido após a remoção das regiões inferiores da imagem. Como pode ser notado, esse novo histograma reproduz de forma visível o contorno dos dedos.

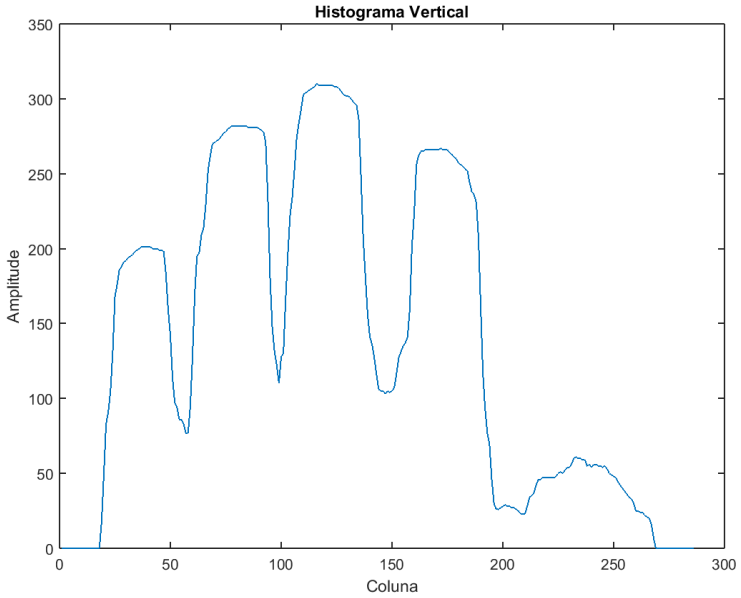


Figura 7. Histograma Vertical da Região dos Dedos

De acordo com a metodologia de Castro, o ponto máximo do histograma vertical está localizado sobre a linha central vertical do dedo médio. A partir desse máximo os primeiros mínimos à direita e à esquerda foram localizados. Esses mínimos delimitam os limites laterais para o corte e segmentação da subimagem do dedo médio.

Em seu trabalho, Castro adotou o método de Tanner & Whitehouse (1969), em que o dedo indicador e o anelar são descartados. Os cortes para segmentação das subimagens dos dedos polegar e mínimo foram implementados de forma análoga.

Diferentemente de Castro, que recortou apenas três dedos (médio, mínimo e polegar), o presente trabalho propôs a segmentação dos cinco dedos da mão. Para a identificação dos pontos de corte do dedo médio foi adotada a mesma técnica utilizada por Castro. A segmentação dos demais dedos foi executada em duas etapas, ambas executadas a partir do histograma vertical. Primeiramente, buscou-se os máximos de cada dedo da seguinte forma: para o dedo indicador buscou-se o primeiro máximo a partir do primeiro mínimo à direita do dedo médio; para o dedo anelar buscou-se o primeiro máximo a partir do primeiro mínimo à esquerda do dedo médio. Em relação ao dedo polegar e ao dedo mínimo, com base em observações realizadas nos histogramas verticais de diversas imagens, considerou-se que seus máximos estão situados a distâncias médias de 15 pixels em relação ao primeiro pixel diferente de zero na imagem binarizada, a partir da extremidade em cada caso.

Após a localização do máximo de cada dedo, e tendo esses pontos como referência, varreduras locais no histograma vertical determinaram os mínimos entre os dedos, indicador e polegar, e anelar e mínimo. Dessa forma os pontos de corte dos cinco dedos foram localizados. A Figura 8 mostra as cinco subimagens que resultaram da segmentação de uma imagem do atlas de Greulich & Pyle.

As subimagens obtidas foram organizadas em um conjunto de imagens no formato “png” que foi utilizado na segunda fase do trabalho.

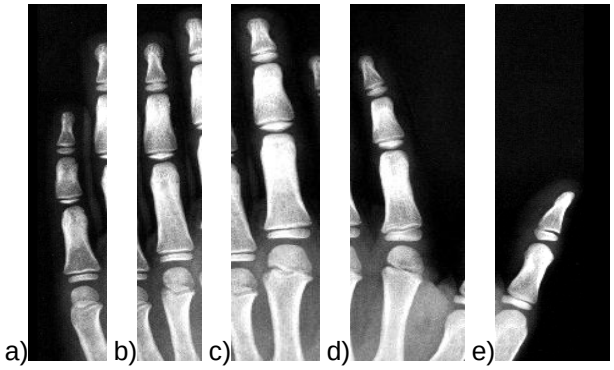


Figura 8. (a) Dedo Mínimo (b) Dedo Anelar (c) Dedo Médio (d) Dedo Indicador (e) Dedo Polegar

2.2. Segunda Fase – Segmentação das Falanges

Como mencionado em seção anterior, o método de segmentação das falanges e epífises utilizado nessa fase, é contribuição original desse trabalho.

Para os testes nessa etapa foi selecionado um conjunto de dez imagens de cada dedo, escolhidas aleatoriamente entre as subimagens obtidas na primeira fase. A condição imposta para a seleção de uma subimagem foi que toda a área de interesse para a segmentação estivesse no quadro da imagem, isto é, que nenhuma parte das falanges estivesse sobre a borda.

Delimitação da Região de Interesse: Como pode ser visto na Figura 9(a), algumas subimagens foram contaminadas com partes dos dedos vizinhos. Dessa forma, foi necessária uma etapa de delimitação da região de interesse para eliminar essas interferências, bem como eventuais ruídos inseridos em etapas anteriores. Deve ser ressaltada a importância da remoção de ruídos para assegurar que somente a região de interesse será segmentada. A remoção de elementos estranhos foi executada sobre a imagem binarizada. Para a binarização foi utilizado o método do limiar, sendo este estabelecido de forma empírica em 140, na faixa entre 0 (zero) e 255. A Figura 9(b) mostra um exemplo de binarização nessa fase.

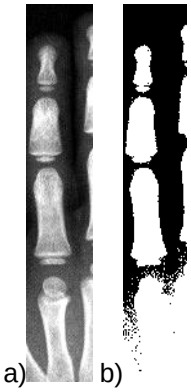


Figura 92. (a) Radiografia do dedo (b) Imagem binarizada

Uma vez binarizada, as regiões pertencentes a dedos vizinhos, bem como as demais estruturas ósseas não pertencentes a região de interesse foram removidas. Deve

ser destacado que nessa fase as regiões de interesse para a segmentação são constituídas pelas falanges proximais, medias e distais. Assim, delimitada a área de interesse, as demais foram eliminadas por um corte horizontal na linha correspondente a vinte pixels abaixo da borda inferior da imagem. A Figura 10 apresenta o resultado do processo de remoção de ruídos.

Figura 30. Imagem Filtrada

Segmentação: De forma geral, a segmentação das estruturas ósseas dos dedos requer uma interferência manual, necessária para apontar a região de interesse, ou a utilização de softwares auxiliares como em Castro [7] e Raymundo [8]. Nesse trabalho é proposto um método autônomo para segmentação das falanges, bem como para a delimitação das epífises, nos casos em que ainda não tenha ocorrido a fusão entre estas partes, o que se dá ao longo do processo de crescimento. Dessa forma, a partir da identificação de pontos de corte horizontais, o algoritmo de segmentação proposto definiu seis partes distintas na estrutura óssea do dedo. O método de busca dos pontos de corte proposto baseou-se na análise dos contornos laterais da estrutura óssea. O procedimento consistiu em realizar varreduras na imagem, primeiramente da direita para a esquerda, para definir o contorno à direita, e em seguida da esquerda para direita, para definir o contorno à esquerda. Em cada caso, a posição do primeiro pixel diferente de zero das linhas foi guardada em um vetor. As Figura 11 e 12 mostram os perfis do contorno à direita e à esquerda, respectivamente. Em cada gráfico, o valor da ordenada, "Posição", indica a posição (coluna) do primeiro pixel diferente de zero da linha correspondente a partir do início da varredura, relativa à imagem da Figura 10.

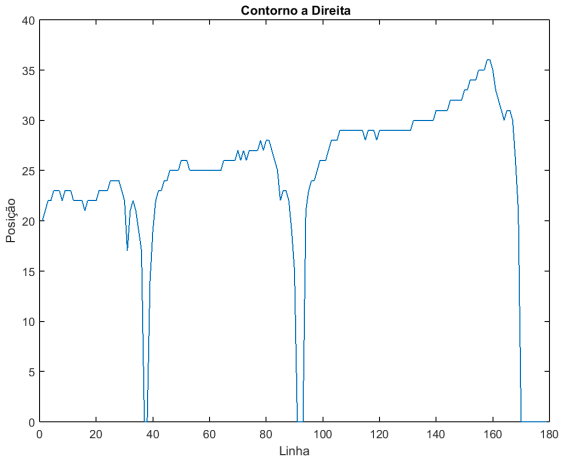


Figura 41. Contorno à Direita

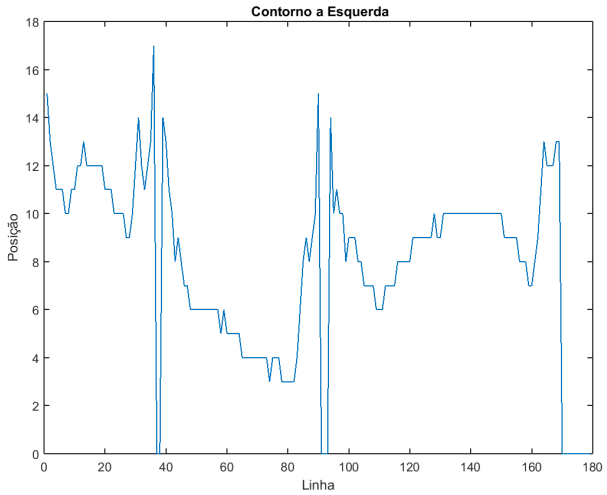


Figura 52. Contorno à Esquerda

Analisando os gráficos dos perfis dos contornos, observa-se que os pontos de corte estão localizados nos vales. Uma varredura nos vetores correspondentes aos contornos elegeu candidatos a pontos de corte, sendo que um ponto será candidato quando a diferença entre ele e um de seus três vizinhos subsequentes e consequentes for maior ou igual a três. Nos casos em haviam mais de um candidato, o escolhido foi sempre o último elemento selecionado. A Figura 13 apresenta um exemplo de resultado da segmentação das falanges pelo método da análise dos contornos laterais.

Figura 63. Imagem do Dedo Segmentada

Avaliação de Desempenho: A avaliação do desempenho consistiu em comparar o resultado final da segmentação com imagens sintéticas de referência que representam a segmentação ideal, geradas manualmente a partir das originais para cada um dos dedos. A Figura 14 mostra dez imagens de referências de dedos específicos, sendo as Figuras 14(a), (b) e (c) relativas ao dedo mínimo; 14(d), (e) e (f), relativas ao dedo anelar; 14(g) ao dedo médio; para o dedo indicador, 14(h) e (i); e para o dedo polegar, 14(j), obtidas a partir do conjunto de imagens dos dedos individuais utilizadas na segunda fase (Segmentação das Falanges). De fato, a avaliação de desempenho verifica a área (número de pixels) efetivamente pertencente à região de interesse que foi inserida na imagem final segmentada.

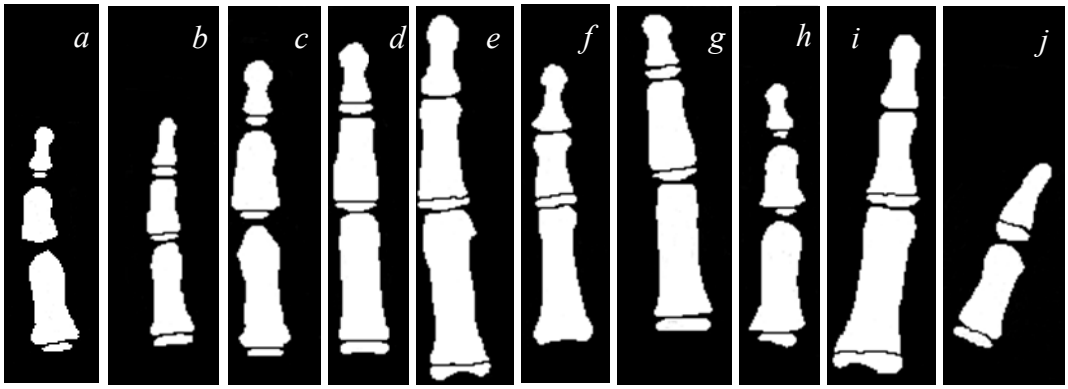


Figura 74. Imagens de referência dos dedos ideais - dedo mínimo: (a), (b) e (c); dedo anelar: (d), (e) e (f); dedo médio: (g); dedo indicador: (h) e (i); e dedo polegar: (j)

Para a avaliação do desempenho, e a fim de normatizar a nomenclatura, designamos como *"osso"* um pixel que pertence à área de interesse, e como *"fundo"* um pixel não pertencente à área de interesse [Moraes 2012]. Nesse caso, quatro situações são possíveis, representadas por VP (Verdadeiro Positivo), VN (Verdadeiro Negativo), FP (Falso Positivo), FN (Falso Negativo), sendo:

VP: o pixel na imagem de referência é osso, e a ele na imagem segmentada é atribuído o status de osso.

VN: o pixel na imagem de referência é fundo, e na imagem segmentada o mesmo recebe o status de fundo.

FP: o pixel na imagem de referência é fundo, mas ele é designado como osso na imagem segmentada.

FN: o pixel na imagem de referência é osso, mas a ele é atribuído o status de fundo na imagem segmentada.

A partir das contagens dos números de VP, VN, FP e FN, as métricas *Precisão*, *Exatidão* e *Erro*, foram calculadas pelas Equações (1), (2) e (3).

$$Precisão = \frac{VP}{VP+FP} \quad \textit{(Equação 1)}$$

$$Exatidão = \frac{VP+VN}{VP+FN+FP+VN} \quad \textit{(Equação 2)}$$

$$Erro = \frac{FP+FN}{VP+FN+FP+VN} \quad \textit{(Equação 3)}$$

Ambiente de Desenvolvimento: Todas as fases do trabalho foram desenvolvidas na plataforma MatLab® R2015a utilizando funções nativas da toolbox de processamento de imagens, além de funções desenvolvidas pelo usuário.

3. Resultados

As Tabelas 1, 2, 3 apresentam os resultados da segmentação. As métricas *Precisão, Exatidão* e *Erro* foram calculadas para os dez dedos, cujas imagens de referência estão representadas na Figura 14. Na Tabela 1 a coluna *Rótulo*, identifica a imagem do dedo

de acordo com sua disposição na Figura 14. As colunas, *Osso Ideal* e *Fundo Ideal* representam as quantidades de pixels correspondentes ao osso e ao fundo nas imagens de referência. As colunas, *Osso Segmentado* e *Fundo Segmentado* representam as quantidades de pixels correspondentes ao osso e ao fundo na imagem segmentada. A coluna *Falso Positivo* indica a quantidade de pixels na imagem de referência que são fundo que foram identificados como osso. A coluna *Falso Negativo* indica a quantidade de pixels nas imagens de referências correspondentes ao osso que foram identificados como fundo.

Tabela 1. Valores de contingências encontrados

Quantidade de pixels								
Rótulo	Osso Ideal	Fundo Ideal	Osso Segmentado	Fundo Segmentado	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo
a	1777	12359	1648	12488	1627	12338	21	150
b	1848	19080	1749	19179	1717	19048	32	131
c	2868	18780	2671	18977	2632	18741	39	236
d	3427	6428	3264	6591	3240	6404	24	187
e	4484	9486	4125	9845	4043	9404	82	441
f	2974	7998	3001	7971	2851	7848	150	123
g	3721	9344	3675	9390	3568	9237	107	153
h	2474	12142	2325	12291	2291	12108	34	183
i	4315	14709	4067	14957	3950	14592	117	365
j	1840	54709	1682	54867	1646	54673	36	194

Baseado nas quantidades de *Verdadeiros Positivos*, *Verdadeiros Negativos*, *Falsos Positivos* e *Falsos Negativos*, as métricas *Precisão*, *Exatidão* e *Erro* foram calculadas, e os resultados estão apresentados na Tabela 2. Pode-se observar que considerando a *Precisão*, o melhor resultado (99,26%) foi obtido para a segmentação da imagem correspondente ao dedo rotulado como *"d"* na Figura 14. Para a Exatidão, o melhor resultado (99,59%) foi obtido para a imagem correspondente ao dedo com rótulo *"j"*. A menor taxa de erro (0,41%) foi obtida para a segmentação do dedo rotulado como *"j"*.

Tabela 2. Valores de métricas encontrados

Métricas (%)			
Rótulo	Precisão	Exatidão	Erro
a	98,73	98,79	1,21
b	98,17	99,22	0,78
c	98,54	98,73	1,27
d	99,26	97,86	2,14
e	98,01	96,26	3,74
f	95,00	97,51	2,49
g	97,09	98,01	1,99
h	98,54	98,52	1,48
i	97,12	97,47	2,53
j	97,86	99,59	0,41

A Tabela 3 apresenta as médias dos resultados obtidos para os dez dedos segmentados considerando cada uma das métricas utilizadas. Como pode ser notado, os valores das métricas obtidos para os demais dedos são comparáveis aos melhores resultados. Dessa forma, considerando que imagens são afetadas de formas diferentes por diferentes fatores, pode-se argumentar que o método de segmentação utilizado possui boa capacidade de generalização.

Tabela 3. Média das Métricas encontradas na Tabela 2

Média das Métricas (%)	
Precisão	97,83
Exatidão	98,20
Erro	1,80

As Figuras 15 e 16 ilustram dois exemplos de resultados da segmentação referentes a dois dedos entre os dez mostrados nas Figura 14. Em cada caso temos, (a) a imagem original; (b) a imagem de referência; (c) a imagem segmentada; (d) os pixels que foram classificados como falsos; (e) os pixels que foram classificados como falsos negativos; e (f) resultado da junção das imagens (c), (d) e (e) sendo que, os pixels referentes aos falsos positivos foram destacados em verde e os falsos negativos em vermelho.

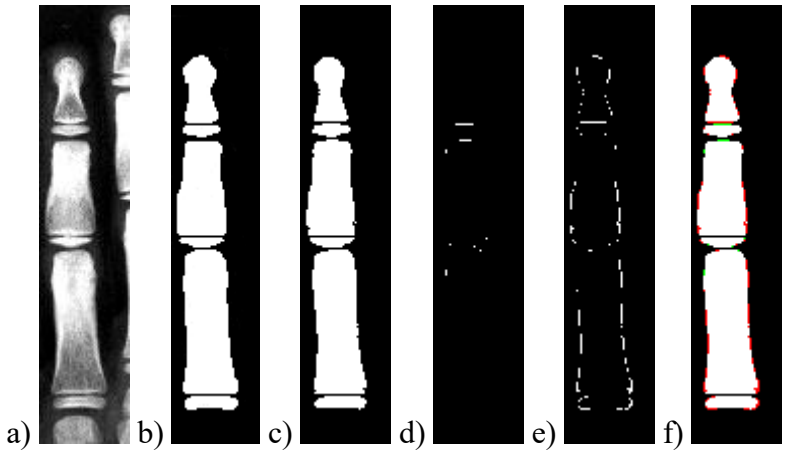


Figura 85. (a) original; (b) ideal; (c) segmentada; (d) falso positivo; (e) falso negativo; (f) analisada (falso positivo em verde e falso negativo em vermelho)

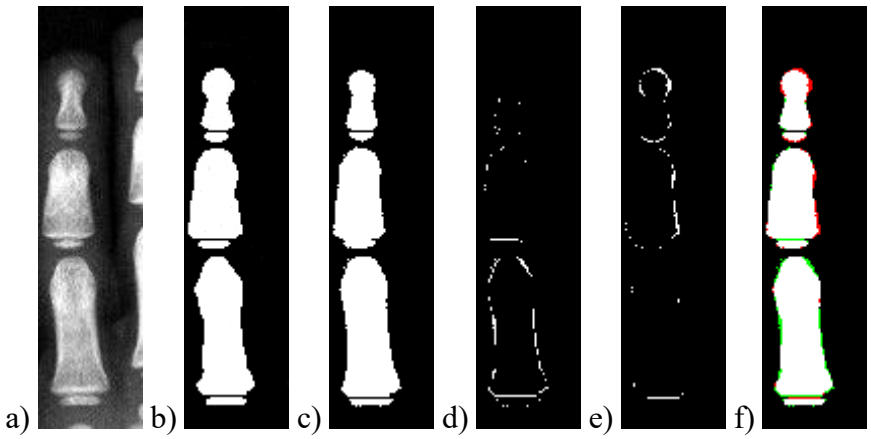


Figura 96. (a) original; (b) ideal; (c) segmentada; (d) falso positivo; (e) falso negativo; (f) analisada (falso positivo em verde e falso negativo em vermelho)

4. Considerações Finais

Nesse trabalho foi apresentada uma metodologia baseada em técnicas de Processamento Digital de Imagens para segmentação das falanges em radiografias carpais. A segunda fase do trabalho é uma proposta dos autores para a segmentação das falanges sem interferência manual, e sem a utilização de softwares auxiliares. Considerando os

desafios envolvidos na generalização do processamento de conjuntos de imagens, os resultados obtidos, ainda em fase inicial das pesquisas, podem ser considerados promissores, mesmo que o método proposto tenha se mostrado sensível a rotação, problema a ser resolvido no curso da pesquisa.

5. Trabalhos Futuros

Pesquisas futuras relacionadas ao tema desse trabalho podem envolver: (1) a investigação de novas técnicas de segmentação, como por exemplo, a utilização da transformada wavelets; (2) a utilização dos resultados da segmentação para o desenvolvimento de algoritmos para extração de características e classificação com vistas a estimação da idade óssea; (3) o desenvolvimento de algoritmos para segmentar outras partes da mão (metacarpais e carpais), de igual importância para a estimação da idade óssea; (4) o aumento do número de imagens do teste para melhorar a avaliação de desempenho da metodologia proposta.

6. Agradecimentos

Os autores desejam agradecer à FAPES pela bolsa da aluna Ding Yih An, pelo apoio do professor Dr. José Geraldo Orlandi, diretor geral do IFES, Campus Serra e da coordenadoria do Mestrado Profissional de Engenharia de Controle e Automação deste mesmo campus.

Referências

Graça, R. F. P. S. O. (2012) “Segmentação de Imagens Torácicas de Raio-X”, 74 f. Dissertação (Mestrado) - Curso de Engenharia Informática, Universidade da Beira Interior, Covilhã.

Greulich , W. W. and Pyle, S. I. (1992) “Radiographic Atlas of Skeletetal Development of the Hand and Wrist”, 2. Ed., Ed. University Press.

Tanner, J. M., Whitehouse R. W. and Healy, M. J. R. (1969) “A New System for Estimating Skeletal Maturity from Hand and Wrist, with Standarts Derived from a Study of 2600 Healthy British Children”, Departament of Growth and Development Institute of Child Health. University of London; and Departament os Statistics, Rothamsted Experimental Station, Harpenden.

Moraes, D. et al, (2012) “Metodologia para Segmentação de ROI Visando a Extração Automática de Características das Vértebras Cervicais para Estimação de Idade Óssea”, VIII Workshop de Visão Computacional, Goiânia.

Castro, F. C. (2009) “Localização, Segmentação e Classificação Automáticas de Regiões de Interesse para a Determinação de Maturidade Óssea Utilizando o Método de Tanner-Whitehouse”, 2009. 120 f. Dissertação (Mestrado) - Curso de Pós-graduação em Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia.

Olivete Jr, C., Rodrigues, E. L. L., & do Nascimento, M. Z. (2005) “O efeito da correção do Efeito Heel em imagens radiográficas da mão,” Revista Brasileira de Física Médica, 1(1).

Behiels, G., Maes, F., Vandermeulen, D. & Suetens, P. (2002) "Retrospective Correction of the Heel Effect in Hand Radiographs", Medical Image Analysis, vol. 6, pp. 183 – 190.

Raymundo, E. M. (2009) "Metodologia de estimação de idade óssea baseada em características métricas utilizando minaradores de dados e classificador neural", 2009. 129 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos.

Santana, G. (2010) "Idade ossea: Atlas de Greulich & Pyle," Disponível em: <http://pt.slideshare.net/endocbh/idade-ossea>. Acesso em: 31 ago. 2016.

Pádua, R. D. S. et al (2008) "Auxílio à detecção de anormalidade perfusional miocárdica utilizando atlas de SPECT e registro de imagens: resultados preliminares," Radiologia Brasileira, v. 41, n. 6, p. 397-402.

Gonçalves, I. S. and Júnior, A. L. A. (2010) "Uma ferramenta baseada em Realidade Aumentada para Visualização de Próteses em Pacientes de Cirurgia," Seminário de Iniciação Científica da Universidade Estadual de Feira de Santana (UEFS).

Macedo, S. O. D. e Oliveira, L. L. G. (2012) "Desenvolvimento de um sistema de auxílio ao diagnóstico de pneumonia na infância utilizando visão computacional," VIII WORKSHOP DE VISÃO COMPUTACIONAL, Goiânia, p. 1 - 8.

Gonzalez, R. C., & Woods, R. E. (2010) "Processamento digital de imagens," tradução: Cristina Yamagami e Leonardo Piamonte.

Souto, L. P. M. (2014) "Mineração de Imagens para a Classificação de Tumores de Mama," 108 f. Dissertação (Mestrado) - Curso de Pós-graduação em Ciência da Computação, Universidade Federal Rural do Semi-Árido, Mossoró.

Soares, H. B. (2008) "Análise e classificação de imagens de lesões da pele por atributos de cor, forma e textura utilizando máquina de vetor de suporte," 180 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Universidade Federal do Rio Grande do Norte, Natal.

Hsieh, C. W. et al (2011) "Fast and fully automatic phalanx segmentation using a grayscale-histogram morphology algorithm," Optical Engineering, 50(8), 087007-087007.

Hsieh, C. et al (2012) "Automatic segmentation of phalanx and epiphyseal/metaphyseal region by gamma parameter enhancement algorithm," Measurement Science Review, 12(1), 21-27.

3D Masks Preprocessing by Direction Vectors applied to Face Recognition

Amauri A. B. Filho[1], Robson da S. Siqueira[1], José M. Soares[2], George A. P. Thé[2]

[1]Department of Computer Science – Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) – Maracanaú, CE – Brazil

[2]Department of Teleinformatics Engineering – Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brazil

amauriairesfilho@gmail.com, siqueira@ifce.edu.br, {marques, george.the}@ufc.br

Abstract. *Depth images have been widely used along with RGB images on the optimization of face segmentation and recognition algorithms. One of the most successful techniques involves the use of facial deformable models in RGB images and the projection, of the 2D masks previously acquired, on depth images. In this work we propose a facial preprocessing algorithm with face detection performed by an optimized deformable model, followed by the use of direction vectors on the correction of the location of the pixels projected onto the depth image, for the acquirement of 3D masks and their use on facial recognition systems. This methodology has shown to be adequate for this task, confirmed by its accuracy, small standard deviation and false positive rate.*

1. Introduction

Facial features extraction, as well as face recognition, has several applications in the entertainment industry, security [Erdogmus and Marcel 2014], [Khan and Updahyay 2014], [Kelkboom, Gokberk and Kevenaar et al. 2007], health care [Amirav, Luder and Halamish et al. 2014], [Lijuan, Zhiyong and Jian et al. 2011], among other realms. Many techniques seek to improve the reliability, efficiency and performance of some already existing algorithms, as mentioned in [Rawlinson, Bhalerao and Wang et al. 2009], [Prabhu and Seshadri 2009], [Scheenstra, Ruifrok and Veltkamp 2005], [Hao, Liao and Qiu 2016], [Hiremath and Hiremath 2013] and [Wang, Hu and Gu 2016]. The utilization of just images in RGB color mode for this task shows to be inefficient in many cases, specially when the images available do not present the predetermined pattern of illumination, shadowing, rotation and colors [Hiremath and Hiremath 2014], [Capor-Hrosik, Vukovic and Pikula 2013], [Zhang and Gao 2009] and [Manolova and Tonchev 2014]. Besides that, 2D images hide characteristics present in the shape of the face. [Zhang 2012] shows that depth sensors allow more interaction of computational systems with the external environment. [Simonite 2014] discusses upcoming smartphones and tablets trends, and explains that even after a decade, the camera is still one of the features most used, and that the implantation of depth cameras will possibly create a new reality in 3D vision in these devices.

In [Baggio, Escrivá and Mahmood 2012], the authors explain that many facial recognition systems work in a proper way when the train and test images are in the same conditions, but their performance is considerably reduced otherwise, characterizing systems very sensitive to the exact conditions in the images. They point out that face preprocessing steps aim to reduce this problem. For example, through the alignment, the features of each sample of the face are located in the same position. With depth images, it is possible to explore facial symmetry through reference points, applying rotations and translations in the set of points that define the face. They also show that a good preprocessing stage increases the reliability of the entire facial recognition system.

Some works have obtained considerable results in the study of the information extracted from depth images along with RGB images. [Ansari, Abdel-Mottaleb and Mahoor 2006] developed a tridimensional method for face modeling, from two frontal and one profile view stereo images, with feature extraction performed by the analysis of disparity maps, applied in 3D facial recognition. The authors describe that through feature extraction in 2D images and the correspondent disparity maps it is possible to calculate their tridimensional coordinates. These 3D features are aligned to a tridimensional mesh model of low resolution and reprojected to the 2D image. This 3D mask can be utilized for systems with high recognition rates. For example, calculating the weighted distances between correspondent vertices in the 3D masks obtained in the work mentioned above, the facial recognition system obtained accuracy of 98%. Another approach is described in [Chang, Bowyer and Flynn 2003], where the authors compare the use of 2D and depth images separately, having Principal Component Analysis (PCA) as main technique for the recognition algorithm, and their use together. An 83.1% accuracy was obtained for 2D images. The utilization of depth images as the input for the facial recognition system resulted in an increase of 0.4% accuracy rate. The use of 2D images along with depth images, where the classification was based on the weighted sum of the distances, presented a superior result, with accuracy of 92.8%.

For the initial process of detecting and segmenting the face, many works have used facial deformable models applied to RGB images. Some optimizations that have been done in deformable models can be observed in [Wang, Ding and Fang 2008], [Zhou, Fang and Li 2014] and [Milborrow and Nicolls 2008].

In this work we propose the development of a face preprocessing method consisting of 5 stages: alignment, detection, projection, correction of the points location by direction vector and the determination of the 3D mask surface area, and its utilization in a facial recognition system. Our methodology makes use of face detectors based on Local Binary Pattern (LBP) and Haar classifiers in order to enhance the initialization of the deformable model fitting algorithm that was optimized in the form of subspace constrained mean-shifts, as it is described in [Saragih, Lucey and Cohn 2009]. In [Zhang, Chu and Xiang 2007] and [Mena, Mayoral and Díaz-Lópe 2015], the authors analyze some variations of Haar and LBP classifiers.

The projection of the facial deformable model in depth images is optimized in this work by direction vectors, having the nose tip as reference for each pixel of the facial model. Thus, we can obtain symmetric 3D masks containing specific features of each individual of the database. In our approach, we calculate the surface areas of the 3D masks and make their use as a preprocessing parameter. For the facial recognition, in [Min, Medioni and Dugelay 2012], they apply the use of Iterative Closest Point (ICP) with multiple canonical faces. We used two approaches for the analysis of the preprocessed data and its use in face recognition. In the first one, we utilized ICP with multiple canonical faces in point clouds with real world dimensions of part 1 of the database. In the second approach, we have a Multilayer Perceptron (MLP) with threshold activation functions applied in the first two parts of the database.

The rest of this article is divided in 4 sections: Section 2 deals with the data acquisition. In section 3 we have the correction of the deformable model projection on the depth images. Section 4 describes the use of 3D masks, preprocessed by our algorithm, in face recognition. It is divided into three subsections 4.1, 4.2 and 4.3. The first two compare, respectively, the information present in 2 and 3 dimensions and our algorithm with other similar 3D algorithms. Subsection 4.3 deals with the use of 3D masks for face recognition performed by a MLP, presenting a study of accuracy and false positive rates in different scenarios in this neural network. In section 5 we have the conclusion and future works.

2. Data acquisition

In order to test our proposed methodology we used the 3DMAD database [Erdogmus and Marcel 2014], which contains 76500 RGB and depth images of 17 individuals of different ages and gender, and it is divided in 3 parts. Both parts 1 and 2 have 25500 samples of frontal faces of 17 individuals, totaling 51000 samples. The data acquisition for part 2 of the database was done two weeks after part 1, with the same 17 classes, and some individuals presenting with some changes in their hair. Faces in parts 1 and 2 of the database have small angular variations in their *x*, *y* and *z* coordinates. In part 3 of the database, the faces of the individuals are covered by a mask, hindering our ability to utilize it in our analysis. Given that the images from the chosen database were captured by a kinect sensor, they are characterized by their low quality and they are not precisely aligned. [Erdogmus and Marcel 2014] explain that the quality of the 3D masks that will be extracted depends considerably on the reconstruction algorithm. For each individual, we took frontal face samples, corresponding to parts 1 and 2 of the database. In [Kramer, Burrus and Echtler 2012], the authors develop an alignment method for RGB and depth images since the raw data acquired through the kinect are not aligned. In this method, each point of the point cloud has an appropriate RGB value associated to it. In our approach, in order to obtain the RGB image aligned to its correspondent depth image, we project the point cloud in a bidimensional space, making the third dimension $z = 0$. To localize 2D points of the face, in the initial stage of the preprocessing algorithm, we utilize the facial deformable model, constituted by 66 points, optimized in the form of subspace constrained mean-shifts described in [Saragih, Lucey and Cohn 2009]. This deformable model delimits 5 regions in the face: left and right eyes, nose, mouth and the more external parts of the face, and it is based in a distribution of landmarks with non-parametric representation, homoscedastic kernel density estimate (KDE) with an isotropic Gaussian kernel (1), maximized through mean shift (2) and incorporated in the model (3). The 2D location of the landmark i^{th} of the point distribution model (PDM) is represented by x_i. Parameters of global scaling, rotation, translation and the set of non-rigid parameters, are denoted respectively by s, R, t and q. The point distribution model has this set of parameters $p = \{s, R, t, q\}$.

$$p(li = aligned|x) \approx \sum_{\mu_i \in \Psi_{x_i^c}} \alpha_{\mu_i}^i N(x; \mu_i, \sigma^2 I), \tag{1}$$

$$x_i^{(\tau+1)} \longleftarrow \sum_{\mu_i \in \Psi_{x_i^c}} \frac{\alpha_{\mu_i}^i N(x_i^\tau; \mu_i, \sigma^2 I)}{\sum_{\mu_i \in \Psi_{y_i^c}} \alpha_y^i N(y_i^\tau; \mu_i, \sigma^2 I)} \mu_i, \tag{2}$$

$$Q(p) = \sum_{i=1}^{n} \parallel x_i - x_i^{(\tau+1)} \parallel^2 . \tag{3}$$

LBP and Haar classifiers were applied, in this order, for the determination of the reduced rectangular area where the deformable model will actuate. Figure 1 shows an example of an RGB input image, depth image, and RGB image after the alignment and obtainment of the 2D points of the face. Notice that now the RGB image is aligned to the depth image, but presents shadows that are incorporated by the infrared depth sensor of the kinect.

3. Correction of the deformable model projection on depth images

With the location of the points of the face in the aligned RGB image, we project them on the depth image and obtain its 3D mask as a point cloud with *x*, *y* and *z* converted to centimeters by the equations (4), (5) and (6), as described in [Kramer, Burrus and Echtler 2012]. The coefficients *dc1* and *dc2* represent, respectively, scale and translation parameters of the conversion. The constants *fxd*, *fyd*, *pxd* and *pyd* are the intrinsic parameters of the camera, respectively, focal length on the x axis, focal length on the *y* axis, coordinate of the principal point on the *x* axis, coordinate of the principal point on

the *y* axis.

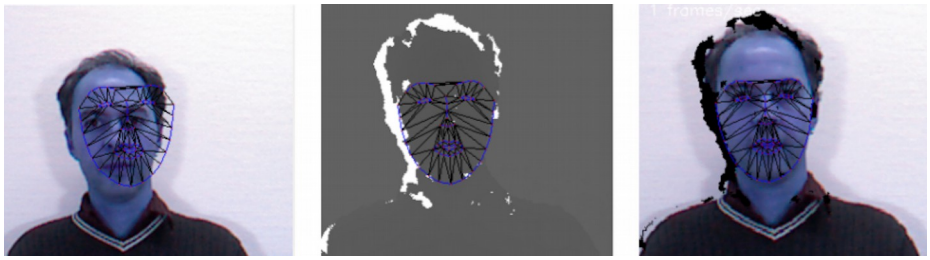


Figure 1. RGB and depth images of individual 1 of the database. On the left is a misaligned RGB image, in the center is the depth image and on the right is the RGB image after the alignment process.

$$z_i = \frac{1}{(depthImage(v_i, u_j) * dc1 + dc2)} * 100, \tag{4}$$

$$x_i = \frac{z_i * (v_j - px_d)}{fx_d}, \tag{5}$$

$$y_i = \frac{z_i * (u_i - py_d)}{fy_d}. \tag{6}$$

Some points of the deformable model projected in the depth image are in shadowing regions that were incorporated by the kinect sensor at the alignment stage of RGB images and their correspondent depth images, described in section II, generating errors as shown in Figure 2.

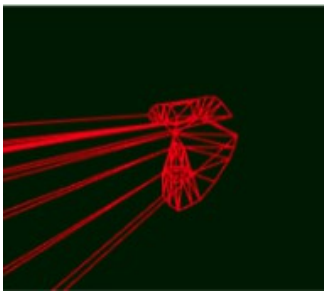


Figure 2. 3D mask without correction.

As explained in the following lines, the correction of the problem above was solved by the use of direction vectors previously defined to each one of the 66 points that constitute the facial deformable model.

Each point p_i localized in shadowing regions moves in the direction of v_i, as we can see in Figure 3 (a), until it finds a region with no shadows. Next, our algorithm calculates the distance between each opposing lateral point of the face and the nose tip, moving again some of these points until the 3D mask becomes approximately symmetric. The algorithm verifies once again if any point belongs to a shadowing region and performs a second correction if necessary. With the 3D mask now approximately symmetric and corrected, we use the 3D point pairs 3 (left side of the face) and 13 (right side of the face) and the points 7 (inferior left side of the chin) and 9 (inferior right side of the chin), respectively, in order to find the angles α and β, rotating the face in the *z* and *y* axes, and translating the face to the center of the coordinates in $p = (0, 0, 0)$, having its center of mass as reference.

$$\alpha = -\arctan(z_3 - z_{13}, x_3 - x_{13}), \tag{7}$$

$$\beta = \arccos \frac{\sqrt[2]{(x_7 - x_9)^2 + (z_7 - z_9)^2}}{\sqrt[2]{(x_7 - x_9)^2 + (y_7 - y_9)^2 + (z_7 - z_9)^2}}. \quad (8)$$

In Figure 3 (b) we see the final stage of a 3D mask approximately symmetric, after the procedure of correction, having all 66 characteristic points, indeed, localized in the face region, which can be verified scanning each of these points. The shadowing regions incorporated by the kinect sensor receive the default value 2047. In Figure 2, among the lines that connect the mask points, there are some with one of their extremities going out of the figure, carrying the default value, and therefore, out of the face region. After the correction is performed, all 66 points have a value different from 2047 and maximum distance from the nose tip of 10 cm in the z coordinate, that is, (p(nose tip).z – $p(i).z$) ≤ 10 cm.

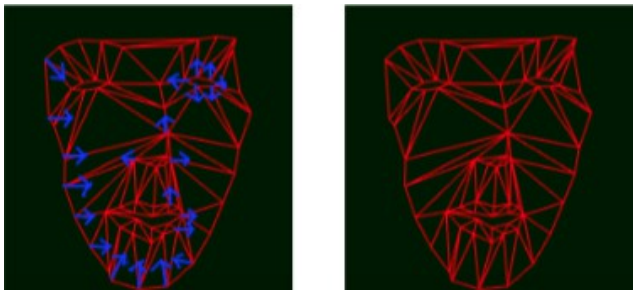


Figure 3. Direction vectors (on the left) and 3D masks corrected (on the right).

With the data converted to centimeters, we calculate the surface area of the 3D mask in cm^2 taking advantage of the triangulation. This information allows us to eliminate detected faces that do not meet the range of surface area (120 - 300 cm^2) of 3D masks of faces analyzed in the initial stage of our preprocessing routine. This range is composed of minimum and maximum values of surface areas obtained from 25500 faces in part 1 of the database. In Figure 4 we can observe the surface areas of the 3D masks of each one of the 17 individuals.

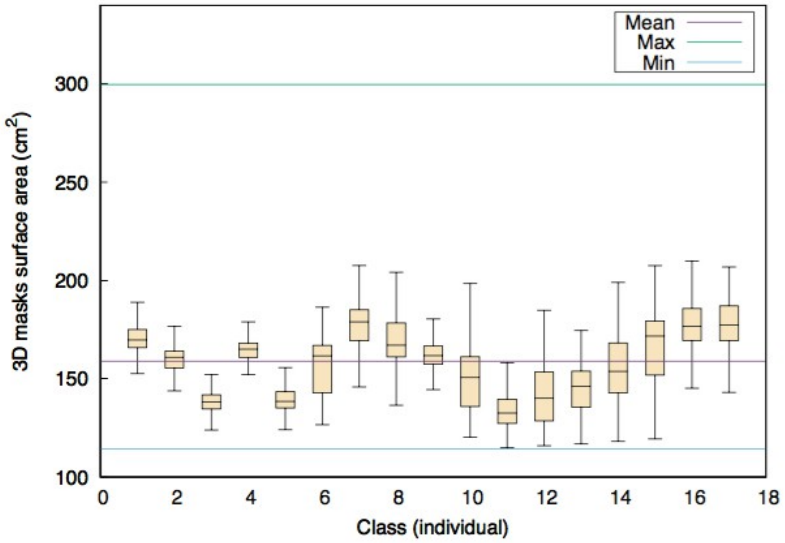


Figure 4. 3D masks surface area for each individual of part 1 of the database.

Our method was able to extract 99.97% of the 3D masks, 50989 out of 51000 samples, where all their points are, indeed, localized in the faces, with an average computational time of 225ms. We conducted our experiments in a consumer-level hardware (Intel(R) Core(TM) i7-3520M CPU @ 2.90GHz). Figure 5 shows a 3D face (a) followed by its triangulation (b) obtained by our method. The algorithm for correction achieved an average computational time of 7ms.

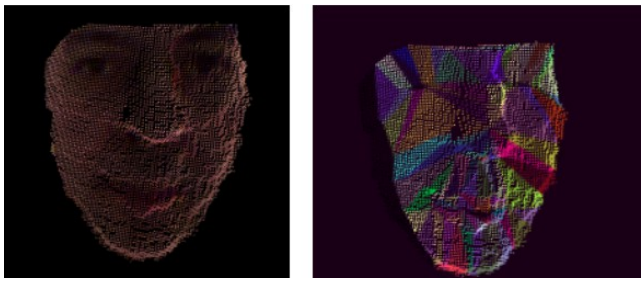


Figure 5. 3D face (on the left) and triangulated face (on the right).

4. Face recognition with preprocessed 3D masks

We utilized two techniques for the analysis of our preprocessing algorithm aiming its use in facial recognition applications. In the first one, the ICP algorithm compares the faces of each individual with correspondent previously chosen model masks, calculating the distance between each of their facial features and assigning the average accuracy as the parameter of comparison between masks with 2 and 3 dimensions. The distance between facial features of individuals has been widely explored by facial recognition algorithms applied in depth images, as mentioned in [Min, Choi and Medioni 2012], [Chang, Bowyer and Flynn 2006] and [Erdogmus and Marcel 2014]. Our second approach utilizes a MLP aiming an increase of accuracy of the classification with low false positive rates. In [Salahshoor and Faez 2012], the authors utilize a MLP for recognition of faces with non-neutral expressions.

4.1. Comparison of the information present in two and three dimensions

Figure 6 and Figure 7 show results of the face recognition performed by the ICP algorithm with multiple canonical faces in 2D and 3D masks. This result was used for the analysis of relevance of the existence of a third dimension in 3D masks. We also used ICP in the 2D masks obtained by the deformable model, assigning the value 0 for the third dimension z_i without correction, and we utilized ICP with the 3D masks preprocessed by our algorithm. Figure 8 is an example of a 2D mask. Both tests were executed with 1, 2, 3, 4 and 5 canonical faces. In all five cases our method showed to be significantly superior.

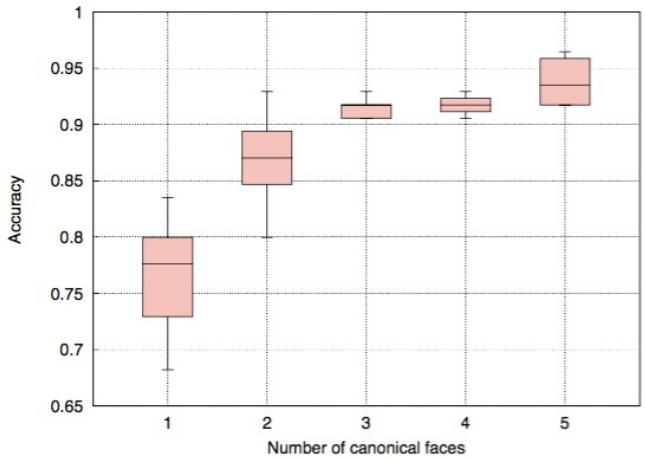


Figure 6. ICP with multiple canonical faces performed in 2D masks.

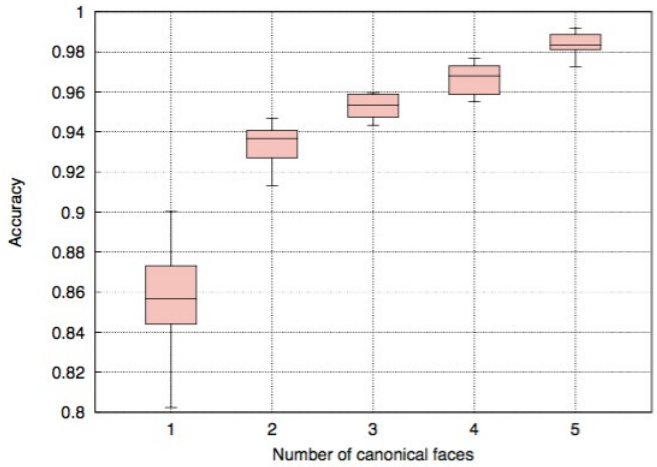


Figure 7. ICP with multiple canonical faces performed in 3D masks.

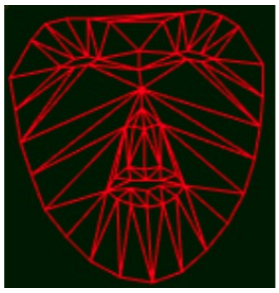


Figure 8. 2D mask of individual 1.

4.2. Comparison between our method and other approaches

In this section we compare different algorithms through the analysis of the following parameters: Nt - Amount of test images, NcflBm - Number of canonical faces or best matches and Ac - Accuracy. Databases with the symbol (-) are unknown. In Table 1 we have the results of the recognition of frontal faces with neutral expressions performed by the algorithms in [Min, Choi and Medioni 2012], [Moreno, Sánchez and Vélez 2003] and our method. With 1, 3 and 4 canonical faces the values of accuracy of our method were slightly inferior to those in [Min, Choi and Medioni 2012], but with the advantage that they were achieved in a test set that is 24 times greater than the one in [Min, Choi and Medioni 2012].

Table 1. Accuracy of different 3D facial recognition methods.

Method used	Database	Nt	Ncf/Bm	Ac (%)
[Min, Choi and Medioni 2012]	-	1054	1	86,82
[Moreno, Sánchez and Vélez 2003]	-	250	1	78,00
Our method	3DMAD	25449	1	**85,95**
[Min, Choi and Medioni 2012]	-	1054	2	91,46
[Moreno, Sánchez and Vélez 2003]	-	250	2	84
Our method	3DMAD	25449	2	**92.83**
[Min, Choi and Medioni 2012]	-	1054	3	96.3
[Moreno, Sánchez and Vélez 2003]	-	250	3	86
Our method	3DMAD	25415	3	**95.45**
[Min, Choi and Medioni 2012]	-	1054	4	97.15
Our method	3DMAD	25449	4	**96.85**
[Min, Choi and Medioni 2012]	-	1054	5	97.91
[Moreno, Sánchez and Vélez 2003]	-	250	5	92
Our method	3DMAD	25381	5	**98.32**

Our algorithm obtained average accuracy of 98.32% with standard deviation of $5.7 \cdot 10^{-3}$ for part 1 of the 3DMAD database with 5 canonical faces. In a real time application, the proposed methodology could achieve a higher face recognition accuracy when analyzing a certain amount of frames per individual instead of the frame by frame analysis, which is confirmed by the low-order standard deviation of 10^{-3} of the average accuracy. The minimum value achieved within the average accuracy of identification with 5 canonical faces was 97.24%.

4.3. MLP for the 3D masks

The methodology adopted in the previous subsection details the superiority of our 3D preprocessing method in comparison to the results obtained with the information in 2D images, however, it does not deal with the fact that the version of the ICP used for the classification stage is a time consuming process.

A MLP was utilized as our second method to verify the relevance of the information present in the 3D masks after the preprocessing stage, showing results that point our methodology as a fundamental stage for classification with depth images. Figure 9 and Figure 10 show results of the 3D masks extraction stage in part 2 of the database. We can see that the faces of both individuals are slightly rotated in relation to the kinect sensor.

Figure 9. Individual 1 of part 2 of the 3DMAD database. Without mask (on the left) and with 2D mask (on the right).

The values of accuracy and false positive of the face recognition performed by the MLP in parts 1 and 2 of the database are represented, respectively, in Figure 11, Figure 12, Figure 13 and Figure 14. The minimum and maximum values of the face recognition occurred, respectively, for the training/testing scenarios 10/90% and 90/10%. For the first part of the dataset, the minimum average accuracy obtained was 99.59%, false positive of $6.85 \cdot 10^{-4}$ and standard deviation of $9.0 \cdot 10^{-4}$. The maximum average accuracy achieved 99.85%, with $5.88 \cdot 10^{-4}$ of false positive and standard deviation of $7.0 \cdot 10^{-5}$. The second part presents 99.71%, $6.53 \cdot 10^{-5}$ and $6.0 \cdot 10^{-6}$, respectively, for minimum average accuracy, false positive and standard deviation. For the maximum average accuracy it was obtained 99.95%, with $7.05 \cdot 10^{-5}$ of false positive and standard deviation of $4.0 \cdot 10^{-4}$.

Figure 10. Individual 6 of part 2 of the 3DMAD database. Without mask (on the left) and with 2D mask (on the right).

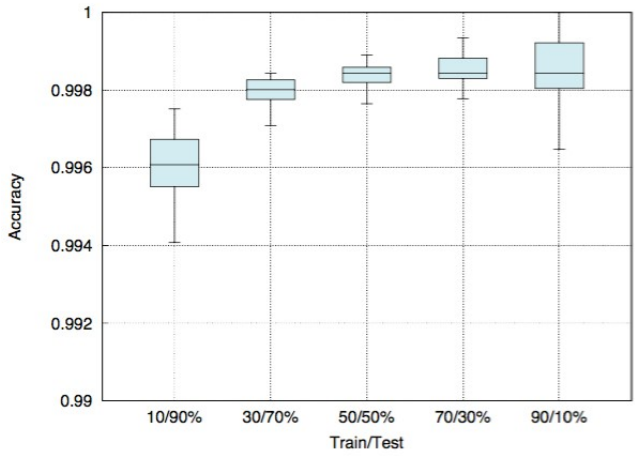


Figure 11. Values of average accuracy for MLP with threshold activation functions applied on the 3D masks of part 1 of the database.

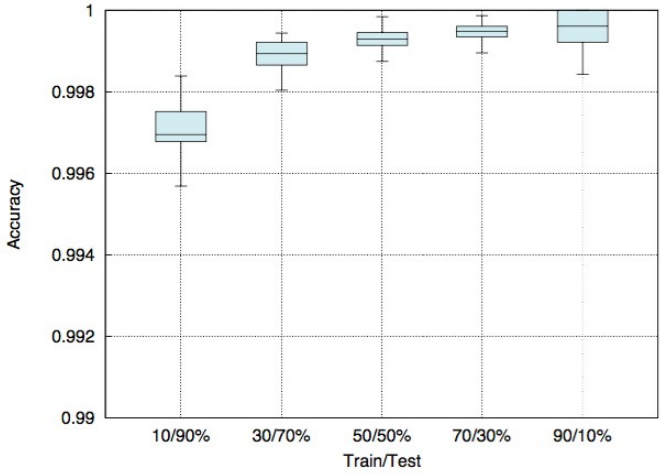


Figure 12. Values of average accuracy for MLP with threshold activation functions applied on the 3D masks of part 2 of the database.

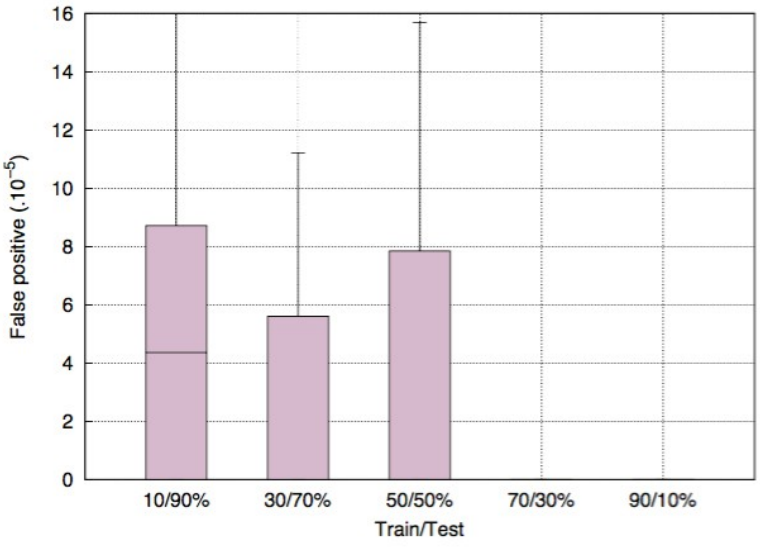


Figure 13. False positive rate of the MLP with threshold activation functions applied on the 3D masks of part 1 of the database.

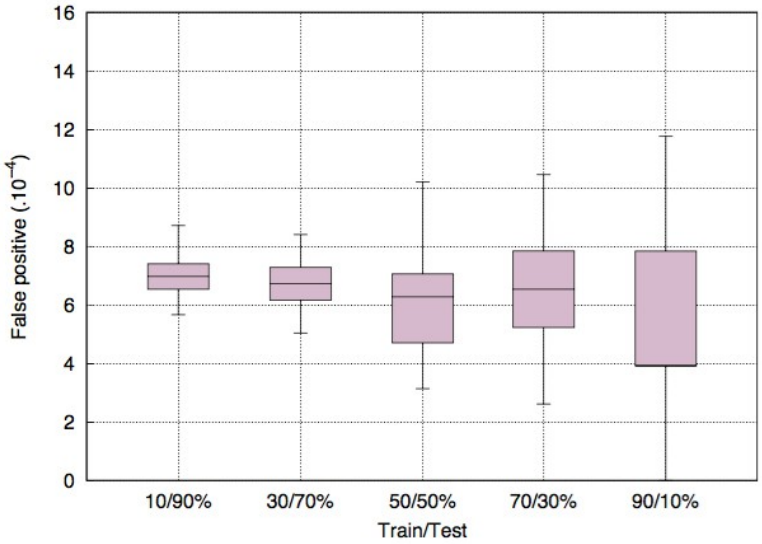


Figure 14. False positive rate of the MLP with threshold activation functions applied on the 3D masks of part 2 of the database.

With the use of a MLP, the computational time was considerably reduced compared to ICP-based approaches for 3D face recognition.

5. Conclusion

In this article, we introduced the development and implementation of a face preprocessing algorithm and its use in face recognition. The classification of the 3D masks obtained through our algorithm, with maximum accuracy of 99.85% and $5.88 \cdot 10^{-4}$ of false positive, indicates high relevance of the extraction of features contained in depth images and the high performance of our preprocessing method with face recognition performed by a MLP. In future works, we plan to analyze more deeply the information of triangulation, texture and superficial area of the faces. We also intend to implant our methodology in a real time facial recognition system, and analyze the performance of our algorithm for correction and extraction in databases that contain faces with different positions of rotations and translations in relation to the depth sensor.

Acknowledgements

The authors would like to thank God, their families and friends, the Universidade Federal do Ceará, the Instituto Federal de Educação, Ciência e Tecnologia do Ceará and the CAPES (Coordination for the Improvement of Higher Education Personnel) for their support.

References

Erdogmus, N., and Marcel, S. (2014). Spoofing face recognition with 3D masks. *IEEE Transactions on Information Forensics and Security*, *9*(7), p.1084-1097.

Khan, J. K., and Upadhyay, D. (2014, September). Security issues in face recognition. In *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference,* pages 719-725. IEEE.

Kelkboom, E. J., Gökberk, B., Kevenaar, T. A., Akkermans, A. H., and van der Veen, M. (2007, August). "3D face": biometric template protection for 3D face recognition. In *International Conference on Biometrics,* pages 566-573. Springer Berlin Heidelberg.

Amirav, I., Luder, A. S., Halamish, A., Raviv, D., Kimmel, R., Waisman, D., and Newhouse, M. T. (2014). Design of aerosol face masks for children using computerized 3D face analysis. *Journal of aerosol medicine and pulmonary drug*

delivery, *27*(4), p.272-278.

Lijuan, S., Zhiyong, A., Jian, Z., Lirong, W., and Qinsheng, D. (2011, August). The 3D Face Feature Extraction and Driving Method on Pronunciation Rehabilitation for Impaired Hearing. In *Computational Science and Engineering (CSE), 2011 IEEE 14th International Conference on,* pages 547-550. IEEE.

Rawlinson, T., Bhalerao, A., and Wang, L. (2009). Principles and methods for face recognition and face modelling. *Handbook of Research on Computational Forensics, Digital Crime, and Investigation: Methods and Solutions: Methods and Solutions*, 53.

Prabhu, U., and Seshadri, K. (2009). Facial Recognition Using Active Shape Models, Local Patches and Support Vector Machines. *contrib. andrew. cmu. Edu*, p.1-8.

Scheenstra, A., Ruifrok, A., and Veltkamp, R. C. (2005, July). A survey of 3D face recognition methods. In *International Conference on Audio-and Video-based Biometric Person Authentication,* pages 891-899. Springer Berlin Heidelberg.

Hao, N., Liao, H., Qiu, Y., and Yang, J. (2016). Face super-resolution reconstruction and recognition using non-local similarity dictionary learning based algorithm. *IEEE/CAA Journal of Automatica Sinica*, *3*(2), p.213-224.

Hiremath, P. S., and Hiremath, M. (2013). 3D Face Recognition based on Deformation Invariant Image using Symbolic LDA. *International Journal*, *2*(2).

Wang, X., Hu, H., and Gu, J. (2016). Pose robust low-resolution face recognition via coupled kernel-based enhanced discriminant analysis. *IEEE/CAA Journal of Automatica Sinica*, *3*(2), p.203-212.

Hiremath, P. and Hiremath, M. (2014). 3D Face Recognition Based on Depth and Intensity Gabor Features using Symbolic PCA and AdaBoost. *International Journal Of Signal Processing, Image Processing And Pattern Recognition*, *6*(5), p.1-12. http://dx.doi.org/10.14257/ijsip.2013.6.5.01

Capor-Hrosik, R., Vukovic, M. T. M., and Pikula, M. (2013). Face Detection by Neural Networks Based on Invariant Moments. In *Mathematical Models for Engineering Science (MMES '13), 4th International Conference on*, p. *45-50*.

Zhang, X., and Gao, Y. (2009). Face recognition across pose: A review. *Pattern Recognition*, *42*(11), p.2876-2896.

Manolova, A., and Tonchev, K. (2014). Study of Two 3D Face Representation Algorithms Using Range Image and Curvature-Based Representations. *International Journal of Computing*, *13*(1), p.42-49.

Zhang, Z. (2012, February). Microsoft kinect sensor and its effect. *IEEE multimedia*, *19*(2), p.4-10.

Simonite, T. (2014). *Laptops and Tablets to Get Depth-Sensing Cameras in 2014 Thanks to Intel. MIT Technology Review*. Retrieved 14 July 2016, from http://www.technologyreview.com/s/523336/ ces- 2014- intels- 3- d- camera- heads- to- laptops- and- tablets.

Baggio, D. L., Escrivá, D. M., Mahmood, N. (2012). *Mastering OpenCV with practical computer vision projects*. Packt Publishing Ltd.

Ansari, A. N., Abdel-Mottaleb, M., and Mahoor, M. H. (2006, July). Disparity-based 3D face modeling using 3D deformable facial mask for 3D face recognition. In *2006 IEEE International Conference on Multimedia and Expo,* pages 981-984. IEEE.

Chang, K. I., Bowyer, K. W., and Flynn, P. J. (2003, October). Multimodal 2D and 3D biometrics for face recognition. In *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on,* pages 187-194. IEEE.

Wang, L., Ding, X., and Fang, C. (2008, July). Generic face alignment using an improved active shape model. In *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on,* pages 317-321. IEEE.

Zhou, L., Fang, B., Li, W., and Peng, L. (2014). Facial feature localization using robust active shape model and POEM descriptors. *Journal of Computers*, *9*(3), p.717-724.

Milborrow, S., and Nicolls, F. (2008, October). Locating facial features with an extended active shape model. In *European conference on computer vision,* pages 504-513. Springer Berlin Heidelberg.

Saragih, J. M., Lucey, S., and Cohn, J. F. (2009, September). Face alignment through subspace constrained mean-shifts. In *2009 IEEE 12th International Conference on Computer Vision,* pages 1034-1041. IEEE.

Zhang, L., Chu, R., Xiang, S., Liao, S., and Li, S. Z. (2007, August). Face detection based on multi-block lbp representation. In *International Conference on Biometrics* pages 11-18. Springer Berlin Heidelberg.

Mena, A. P., Mayoral, M. B., and Díaz-Lópe, E. (2015, June). Comparative study of the features used by algorithms based on viola and jones face detection algorithm. In *International Work-Conference on the Interplay Between Natural and Artificial Computation,* pages 175-183. Springer International Publishing.

Min, R., Choi, J., Medioni, G., and Dugelay, J. L. (2012, November). Real-time 3D face identification from a depth camera. In *Pattern Recognition (ICPR), 2012 21st International Conference on,* pages 1739-1742. IEEE.

Kramer, J., Burrus, N., Echtler, F., Daniel, H. C., and Parker, M. (2012). *Hacking the Kinect* (Vol. 268). New York, NY, USA:: Apress.

Chang, K. I., Bowyer, K. W., and Flynn, P. J. (2006). Multiple nose region matching for 3D face recognition under varying facial expression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(10), p.1695-1700.

Salahshoor, S., and Faez, K. (2012, June). 3D face recognition using an expression insensitive dynamic mask. In *International Conference on Image and Signal Processing* pages 253-260. Springer Berlin Heidelberg.

Moreno, A. B., Sánchez, A., Vélez, J. F., and Díaz, F. J. (2003, September). Face recognition using 3D surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference* (Vol. 2003).

Implementação de um Modelo de Simulação de Sistemas Fotovoltaicos Não-Controlados Conectados à Rede de Distribuição de Energia Elétrica

Henrique P. Corrêa[1], Ricardo A.P. Franco[1], Flávio H.T. Vieira[1], Marcelo S. de Castro[1]

[1]Universidade Federal de Goiás
Goiânia – GO – Brazil

Resumo. *Neste trabalho é proposto um modelo de simulação, implementado via software MATLAB, de um sistema de geração fotovoltaica não-controlada conectado à rede elétrica de distribuição. O modelo considera fatores pertinentes ao desempenho do sistema fotovoltaico, a saber: os efeitos surtidos sobre a potência gerada pelas condições ambientais (irradiância e temperatura de célula) e o impacto da geração fotovoltaica sobre a rede, no que se refere aos níveis de tensão, introdução de componentes harmônicas de corrente e tensão no sistema e suprimento de potência pela rede elétrica.*

Abstract. *In this work, we propose a MATLAB-based simulation model for uncontrolled photovoltaic generation systems connected to a power distribution grid. The proposed model takes into account the main factors related to photovoltaic system performance: variations in supplied power effected by the change of environmental conditions (irradiance and cell temperature) and the perceived impact of photovoltaic generation upon the grid in regards to voltage levels, introduction of current and voltage harmonic components and power supply by the grid.*

1. Introdução

A situação de geração fotovoltaica contemplada no presente trabalho consiste em uma carga residencial dotada de geração fotovoltaica não-controlada e conectada à concessionária fornecedora de potência trifásica, a qual supre outras cargas, supostas isentas de geração distribuída.

Ademais, a simulação de geração fotovoltaica interligada à rede permite a análise do impacto sobre esta, que pode ter a qualidade da energia elétrica fornecida aos consumidores avariada. O que promove a degradação da qualidade de energia, por parte da geração fotovoltaica, é o caráter não-linear deste processo, o qual implica na introdução de componentes harmônicas de tensão e corrente na rede (Mohan, 2003).

A potência e qualidade de energia fornecidas por um sistema fotovoltaico dependem do número e características dos módulos fotovoltaicos que o compõem, do projeto dos equipamentos de interface do sistema (e.g., inversor e filtro de saída) com a rede e de condições ambientais como temperatura e irradiância. O modelo proposto neste trabalho contempla todos estes fatores e proporciona um bloco funcional que pode ser conectado a uma rede arbitrária para a realização de simulações.

Desse modo, o objetivo deste trabalho é contribuir com um modelo de sistema fotovoltaico que possa ser imediatamente conectado a redes trifásicas arbitrárias a serem

simuladas no pacote *SimPowerSystems* do *software MATLAB* e contemple a possibilidade de variações dinâmicas de carga e condições ambientais durante as simulações da rede. O modelo também é passível de ser utilizado em maior escala (vários blocos), o que corresponderia à simulação de redes com presença intensa de geração distribuída.

2. A Célula Fotovoltaica

O componente fundamental dos sistemas fotovoltaicos é a célula fotovoltaica. Trata-se de material semicondutor dopado similarmente a um diodo, caracterizando-se como uma junção pn. Sabe-se que, em uma junção pn iluminada, há tendência simultânea à geração e condução como diodo. Tais considerações, assim como outras acerca de perdas associadas à qualidade do material semicondutor e ao processo de geração fotovoltaica, podem ser modeladas pelo seguinte circuito equivalente (Figura 1) (Olindo et al, 2016):

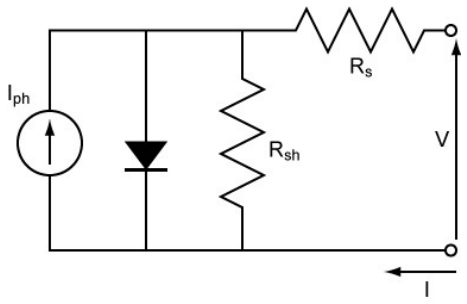


Figura 1. Circuito equivalente da célula fotovoltaica.

No presente estudo, adotamos um modelo simplificado, relativamente ao circuito da Figura 1, em que considera-se $R_{sh} = 0$. Desse modo, pode-se mostrar que a célula possuirá a seguinte característica I-V (Olindo et al, 2016):

$$V = \frac{kT_c}{q} \cdot ln\left(\frac{I_{ph} + I_o - I}{I_o}\right) - R_s I \tag{1}$$

Na Equação (1), os parâmetros k, q e I_o são constantes e referem-se, respectivamente, à constante de Boltzmann, à carga elementar e à corrente de polarização reversa da junção pn em consideração. Embora T_c, a temperatura da célula, seja variável, esta será tratada como constante em (1). Sua variação é contabilizada, nas simulações, conforme explanado no parágrafo seguinte.

Para o estudo da célula fotovoltaica sob condições ambientais variáveis, a Equação (1) deve ser modificada por fatores corretivos que reflitam tais flutuações. A variação de temperatura ambiente e irradiância têm efeitos importantes sobre a fotocorrente I_{ph} e a tensão de saída da célula, V. As equações que modelam tais considerações são (Aramizu, 2010):

$$C_V = [1 + \beta_T(T_c - T)]\,[1 + \beta_T \alpha_S(S - S_c)] \tag{2}$$

$$C_{I_{ph}} = \left[1 + \frac{\gamma_T}{S_c}(T - T_c)\right]\left[1 + \frac{1}{S_c}(S - S_c)\right] \tag{3}$$

Os valores numéricos de C_V e $C_{I_{ph}}$ refletem os desvios das grandezas V e I_{ph}, decorrentes dos parâmetros temperatura ambiente e irradiância (T e S), com relação aos valores que seriam encontrados em condições de referência (T_c e S_c). As constantes α_S, β_T e γ_T são dependentes do projeto da célula fotovoltaica e de condições ambientais diversas das anteriormente citadas, como, por exemplo, a velocidade do vento (Olindo et al, 2016). A tensão terminal e a corrente fotovoltaica efetivamente verificadas são dadas por:

$$V = C_V \cdot V_b \tag{4}$$

$$I_{ph} = C_{I_{ph}} \cdot I_{ph,b} \tag{5}$$

Os parâmetros V_b e $I_{ph,b}$ correspondem aos valores calculados por meio da Equação (1), considerando-se operação em condições ambientais de referência.

Tendo-se em vista este modelo matemático da célula fotovoltaica, a Figura 2 apresenta um diagrama de blocos implementado no *software* MATLAB. Tal diagrama foi usado para modelar-se a célula fotovoltaica segundo as Equações (1) a (5).

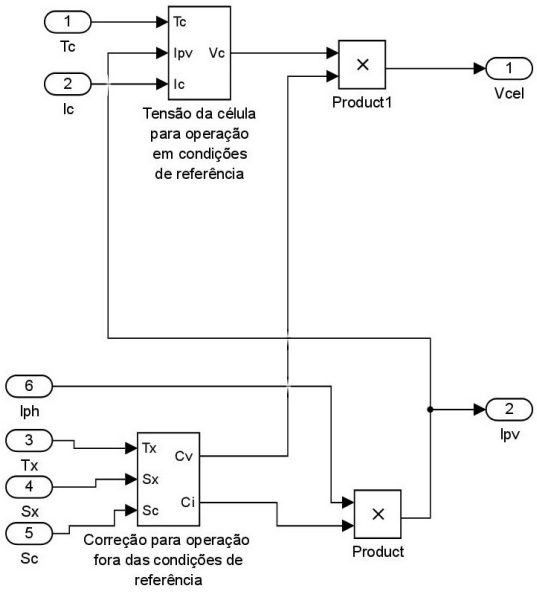


Figura 2. Modelo da célula fotovoltaica.

3. Módulos e Painéis Fotovoltaicos

O módulo fotovoltaico é a unidade usualmente oferecida ao consumidor interessado em tecnologias de geração fotovoltaica. Trata-se de uma combinação específica de células fotovoltaicas, as quais podem vir a ser associadas em série ou paralelo, de modo a atender, nominalmente, uma determinada solicitação de tensão e corrente de carga.

A combinação em série (*string*) de células fotovoltaicas promove um aumento da tensão terminal suprida pelo módulo fotovoltaico. Supondo a associação em série de N

células cujas tensões nominais sejam iguais a V_{cel}, tem-se que a tensão de saída da *string* será de NV_{cel}. Considerações análogas levam à conclusão que M células associadas em paralelo suprirão uma corrente nominal MI_{cel}, em que I_{cel} é a corrente provida por uma célula fotovoltaica individual.

Apresenta-se, na Figura 3, o modelo implementado para um painel fotovoltaico, o qual toma por base o modelo de célula fotovoltaica apresentado na Seção 2 e simula o efeito, sobre V e I, da associação em paralelo de N_p strings contendo, cada uma, N_s células.

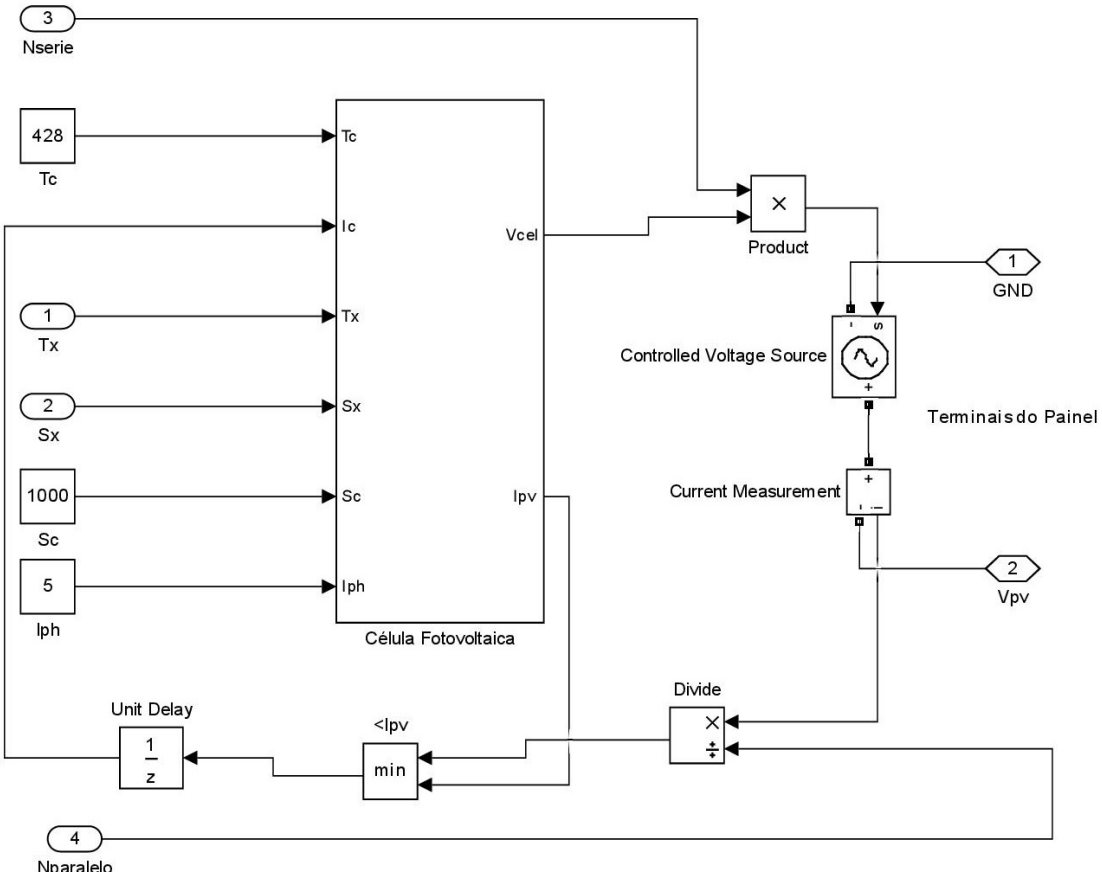


Figura 3. Modelo do painel fotovoltaico.

4. Inversor de Potência e Filtro LC de Saída

Tendo em vista que o painel fotovoltaico opera fornecendo corrente contínua, é claro que este não pode ser ligado diretamente à rede elétrica. Uma interface composta de filtros e equipamentos característicos da eletrônica de potência deve ser utilizada para o processamento da potência entregue pelo painel, convertendo as formas de onda DC de entrada em sinais AC com a menor distorção harmônica total (THD) possível, sendo esta última exigência necessária à manutenção da qualidade de energia na rede elétrica. O cálculo da THD de uma forma de onda periódica $x(t)$ se dá pela Equação 6:

$$THD_x = \frac{1}{X_1}\sqrt{\sum_{n=2}^{\infty} X_n^2} \tag{6}$$

Nesta equação, X_1 refere-se à amplitude da componente fundamental da onda, obtida por meio de sua expansão em série de Fourier. Analogamente, X_n corresponde à amplitude da n-ésima harmônica. Desse modo, vê-se que o parâmetro THD é uma medida do conteúdo harmônico do sinal $x(t)$.

O inversor de potência serve o propósito de processar a potência DC advinda do painel fotovoltaico e transformá-la em potência AC trifásica. Isso é feito por meio de um circuito com elementos de chaveamento controlado cuja entrada é ligada ao painel e saída consiste em três terminais a serem conectados a um filtro LC passa-baixas.

A operação escolhida para o inversor foi a de chaveamento *PWM*, que tem a vantagem de propiciar formas de onda, na saída, com DHT relativamente reduzida (Mohan, 2003). A saída do inversor é conectada a um filtro LC trifásico, o qual atenua o conteúdo harmônico do sinal PWM, de frequência fundamental igual a 60 Hz. Como o filtro não é ideal, obtém-se tensões e correntes com distorção harmônica, avariando-se a qualidade da energia elétrica. A implementação computacional do sistema de interface descrito na presente seção, com o detalhe de conexão à carga residencial, encontra-se na Figura 4.

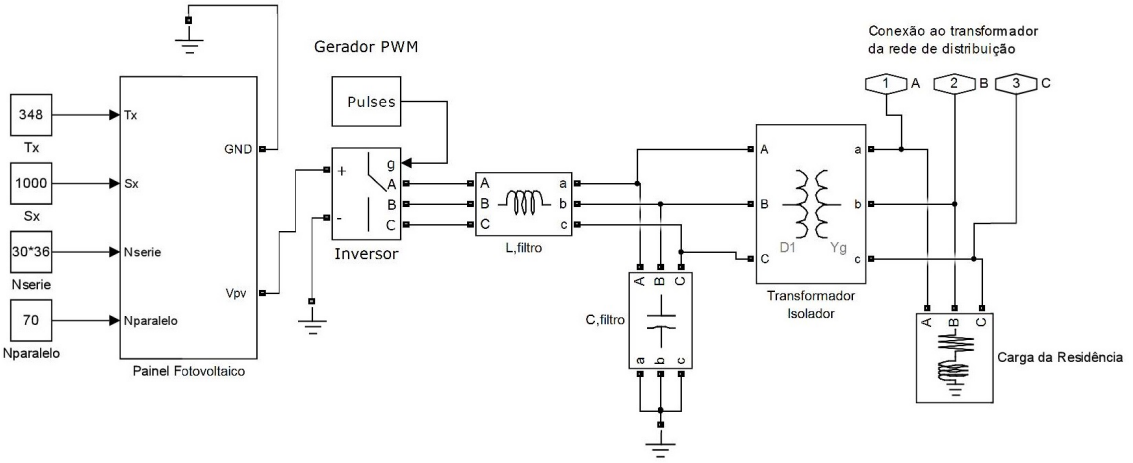


Figura 4. Modelo do sistema fotovoltaico e carga residencial.

5. Simulação e Resultados Obtidos

A simulação realizada neste trabalho consistiu em conectar o modelo de sistema fotovoltaico previamente descrito a uma carga residencial (conforme Figura 4), sendo esta, por sua vez, ligada a uma rede de distribuição alimentada por uma subestação e à qual encontram-se conectadas outras cargas de porte similar à da residência. O esquemático da implementação computacional da rede e parâmetros de simulação utilizados encontram-se na Figura 7, localizada no Anexo.

Observe-se que foram analisadas simulações em que as condições ambientais foram mantidas constantes durante a operação do sistema fotovoltaico. Optou-se por escolher, para a simulação inicial, os valores de temperatura ambiente e irradiância como sendo iguais aos valores de referência do módulo fotovoltaico e uma frequência de chaveamento no inversor igual a 8kHz. Outras simulações, executadas subsequentemente, visaram a análise dos efeitos surtidos sob o sistema devido à mudança desses mesmos três

parâmetros: temperatura ambiente, irradiância e frequência de chaveamento do inversor. Os valores de referência assumidos para o painel foram $T_c = 25°C$ e $S_c = 1000W/m^2$.

5.1. Alimentação Exclusivamente via Concessionária *versus* Geração Fotovoltaica

Simulações foram inicialmente realizadas visando-se a análise e comparação de duas situações distintas, a saber: alimentação da residência exclusivamente pela concessionária (sistema fotovoltaico desconectado) e geração fotovoltaica em condições ambientais de referência (25°C, $1000W/m^2$) e com frequência de chaveamento de inversor igual a 8kHz. Os resultados permitem visualizar impactos sobre a qualidade de energia advindos do sistema fotovoltaico e quanta economia de energia tal sistema seria capaz de prover ao usuário.

De ambas as simulações, foram extraídos os valores *rms* e de THD das formas de onda de tensão na carga (v_{carga}), corrente de saída do sistema fotovoltaico ($i_{sistema}$), corrente drenada do sistema de distribuição (i_{rede}) e tensão nos pontos 4 e 5 da rede (v_4 e v_5, Figura 7). Ademais, foram também aferidas as potências ativa e reativa providas pelo sistema fotovoltaico e absorvidas pela carga residencial; essas últimas consistem em potência provida pela rede e pelo sistema fotovoltaico. Tais resultados encontram-se delineados nas Tabelas 1 e 2.

Tabela 1. Alimentação via concessionária e geração em condições de referência

Grandeza	Concessionária		Microgeração	
(V/A)	RMS	THD(%)	RMS	THD(%)
v_{carga}	197,7	0	199,3	2,6
$i_{sistema}$	0	0	252	3,2
i_{rede}	471,5	0	272	2,0
i_{carga}	471,5	0	474,4	0,6
v_4	$11{,}1 \cdot 10^3$	0	$11{,}2 \cdot 10^3$	1,1
v_5	$11{,}6 \cdot 10^3$	0	$11{,}6 \cdot 10^3$	0,6

Tabela 2. Alimentação via concessionária e geração em condições de referência

Grandeza (kW/kVar)	Concessionária	Microgeração
P_{carga}	145	147
Q_{carga}	70	71
$P_{sistema}$	0	87
$Q_{sistema}$	0	-11

Quando se trata da análise de distorção harmônica na rede elétrica, a forma de onda cuja THD é usualmente utilizada como referência para julgar-se a qualidade de energia provida é aquela da corrente que flui pela rede, cuja distorção advém do elemento não-linear conectado que, neste caso, trata-se do painel fotovoltaico. Ademais, é de especial interesse a distorção da tensão de carga, pois esta também demonstra a capacidade do sistema fotovoltaico de injetar harmônicos na rede (Mohan, 2003). Desse modo, apresentamos, nas Figuras 5 e 6, os gráficos obtidos para estas duas formas de onda.

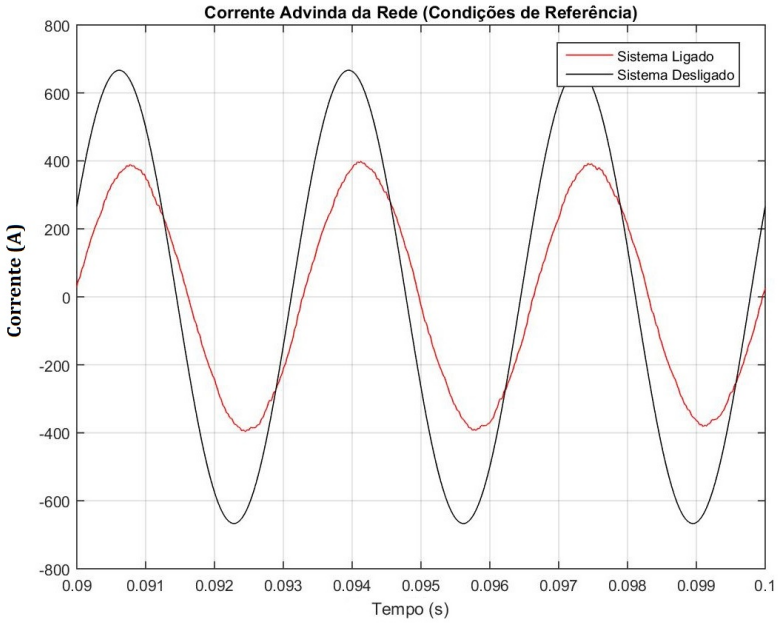


Figura 5. Formas de onda da corrente advinda da rede de distribuição.

5.2. Geração Fotovoltaica em Diferentes Condições Ambientais

Obtidos os resultados referentes à geração em condições de referência, procedeu-se à verificação dos efeitos das condições ambientais (temperatura ambiente, irradiância) e da frequência de chaveamento do inversor sobre a geração. Simulações do sistema foram realizadas para variações individuais de cada um desses parâmetros.

A menos das condições de referência, os parâmetros contemplados foram: temperaturas de 15°C e 35°C, irradiâncias de $600W/m^2$ e $1300W/m^2$ e frequências de chaveamento de 1kHz e 4kHz. Neste caso, julgou-se suficiente a medida de valores *rms* e de THD dos mesmos parâmetros da Tabela 1 da Seção 5.1, com exceção de v_4 e v_5. Os resultados encontram-se nas Tabelas 3 a 5.

Tabela 3. Operação em temperaturas distintas da referência (25°C)

Grandeza	15°C		35°C	
(V/A)	RMS	THD(%)	RMS	THD(%)
v_{carga}	199,7	2,7	199,0	2,5
$i_{sistema}$	258	3,0	244	2,8
i_{rede}	264	2,6	278	1,9
i_{carga}	475,0	0,8	473,6	0,6

5.3. Análise de Resultados e Conclusões

Pela análise das simulações realizadas, pode-se concluir que sistemas de geração fotovoltaica têm o potencial de promover grande economia de energia por parte do consumidor. De modo mais específico, em operação nas condições de referência, de toda a potência consumida pela carga, observa-se (Tabela 2) que 59,2% adveio do sistema fotovoltaico. Considerando-se que o sistema simulado é não-controlado, a possibilidade de economia

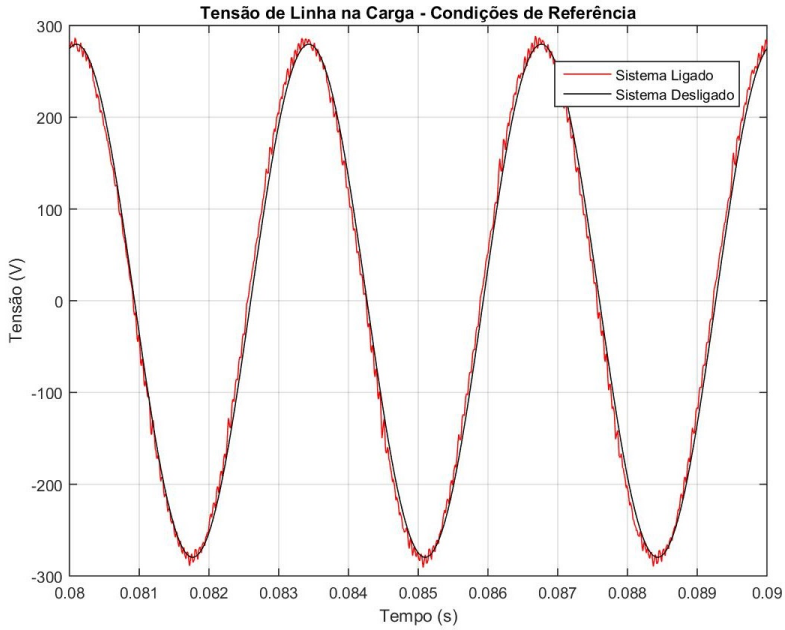


Figura 6. Formas de onda da tensão sobre a carga.

Tabela 4. Operação em irradiâncias distintas da referência ($1000W/m^2$)

Grandeza	$600W/m^2$		$1300W/m^2$	
(V/A)	RMS	THD(%)	RMS	THD(%)
v_{carga}	195,0	1,7	202,3	3,2
$i_{sistema}$	148,0	3	330,0	2,8
i_{rede}	393,5	1,2	193,0	4,5
i_{carga}	464,5	0,6	481,2	0,7

via implementação de sistemas de controle (Olindo et al, 2016) é, potencialmente, ainda maior.

Vê-se que a potência total consumida pela carga não é significativamente alterada pela presença de microgeração na residência. Isto está diretamente ligado à incapacidade do sistema fotovoltaico de interferir grandemente sobre a tensão de carga, o que reflete o fato da subestação aproximar-se de um barramento infinito do ponto de vista da carga. Entretanto, como de fato há impedância entre carga e subestação, percebe-se distorção harmônica na tensão de carga, devido ao fluxo de correntes harmônicas ao longo das linhas de transmissão da rede de distribuição.

Tem-se que a temperatura provoca variações perceptíveis na corrente de saída. Mais especificamente, verifica-se que, quanto menor a temperatura de operação, maior será a corrente de saída e, portanto, a potência suprida. Isto está associado ao decréscimo da taxa de recombinação de portadores no corpo da célula fotovoltaica, quando a temperatura de operação desta é reduzida.

A frequência de chaveamento do inversor mostrou-se determinante para a qualidade de energia elétrica suprida pelo sistema fotovoltaico. Quanto menor a frequência

Tabela 5. Operação em chaveamentos distintos da referência (8kHz)

Grandeza	1kHz		4kHz	
(V/A)	RMS	THD(%)	RMS	THD(%)
v_{carga}	198,0	6,3	199,1	3,1
$i_{sistema}$	244,0	5,4	252,1	4,0
i_{rede}	296,2	3,5	276,3	3,6
i_{carga}	471,0	0,8	473,7	0,7

de chaveamento, maiores os valores de DHT verificados na tensão de carga e corrente de saída do sistema fotovoltaico. Isto explica-se pelo fato de maiores frequências de chaveamento fornecerem correntes, na saída do sistema fotovoltaico, com menor conteúdo harmônico. Como foi utilizado chaveamento PWM no inversor, valores de THD muito elevados (por exemplo, $> 10\%$) não foram verificados, mesmo na operação a 1kHz.

A irradiância, naturalmente, determina a potência entregue pelo sistema fotovoltaico; vê-se que o valor da irradiância influencia fortemente a magnitude da corrente e, portanto, da potência suprida pelo painel fotovoltaico. Desse modo, os resultados confirmam que a performance do sistema está fortemente atrelada à sua instalação em uma área com níveis de irradiância elevados que se prolonguem pelo maior tempo possível.

Neste trabalho, foi considerada a simulação de um sistema fotovoltaico hipotético. Este sistema (Figura 7) possui um total de 75600 células fotovoltaicas, com $I_{ph} = I_{sc} = 5A$ e $V_{(I=0)} = V_{oc} = 0,6V$. Desse modo, a potência máxima que pode ser fornecida pelo sistema é $P_m = 75600 \cdot \eta \cdot V_{oc}I_{sc} \Rightarrow P_m = \eta \cdot 226,8kW$, em que η é o fator de preenchimento da célula, cujo valor usual para células com baixo R_s é 0,7 (Olindo et al, 2016). Desse modo, pode-se estimar $P_m = 158,8kW$. Ademais, para condições ambientais de referência, obtivemos $P_{sistema} = 71kW$. Como a impedância da rede é relativamente elevada, sabe-se que o painel estará operando em um ponto de baixa corrente da característica I-V. Logo, a tensão de saída estará muito próxima de V_{oc} e conclui-se que $I \approx \frac{71kW}{158,8kW} N_p I_{ph} \Rightarrow I \approx 156,5A$. Entretanto, o valor obtido em simulação para I foi de 153,8A, o que demonstra a coerência dos resultados obtidos. Em trabalhos futuros, sistemas reais serão simulados de modo a poderem ser realizadas comparações diretas com o modelo proposto.

Referências

1. Aramizu, J. (2010). Modelagem e Análise de Desempenho de um Sistema Fotovoltaico em Operação Isolada e em Paralelo com uma Rede de Distribuição de Energia Elétrica (Dissertação de conclusão de curso).

2. Wurfel, P. e Olindo, I. (2016). Solar Energy: The Physics and Engineering of Photovoltaic Conversion, Technologies and Systems. UIT Cambridge LTD, 1st Edition.

3. Mohan, N. (2003). Power Electronics: Converters, Applications and Design. John Wiley and Sons, Third Edition.

6. Anexo

Na página seguinte, encontra-se a representação da rede elétrica simulada, conjuntamente com parâmetros pertinentes de simulação.

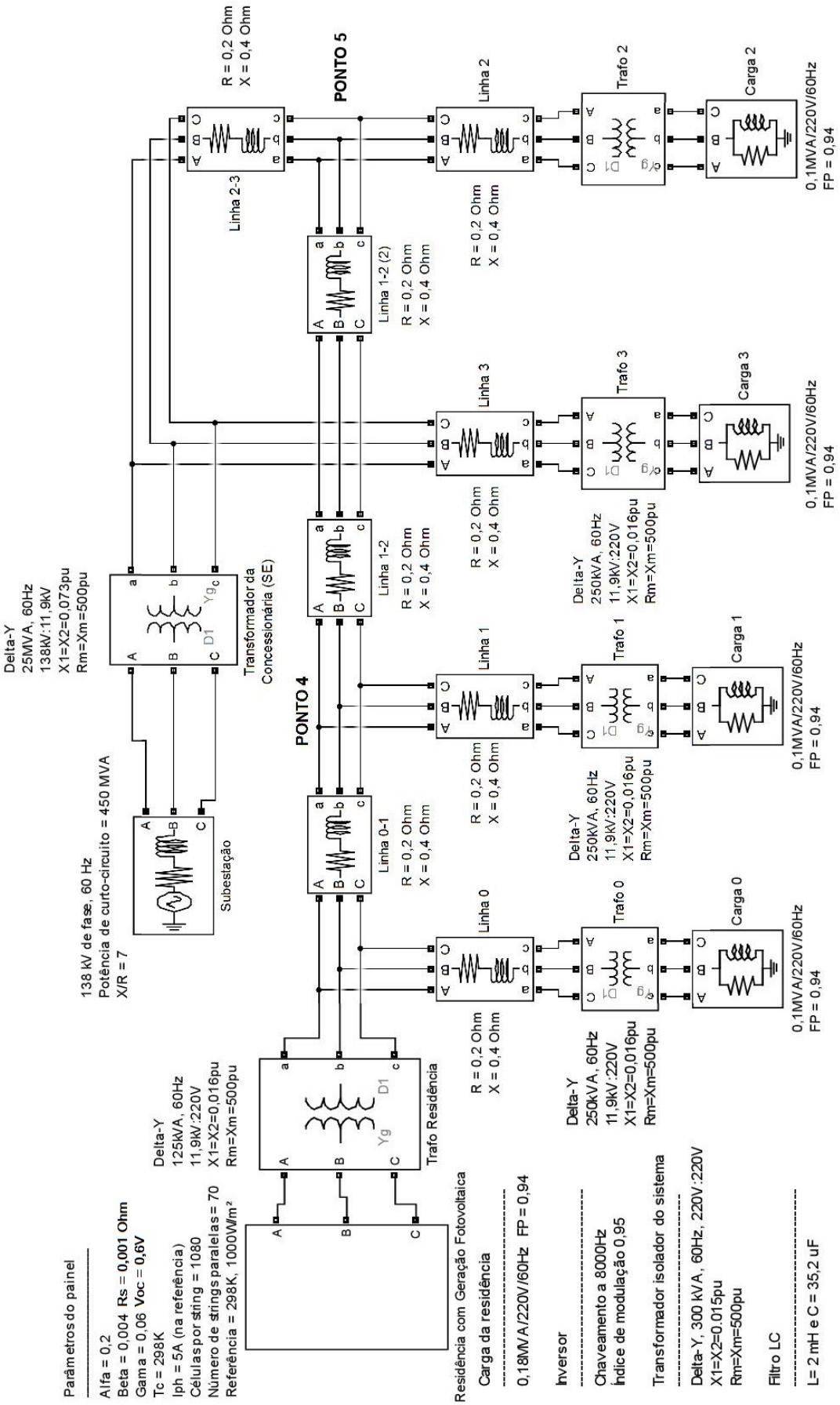


Figura 7. Modelo da residência e da rede elétrica à qual esta se encontra ligada.

Aplicação do Software GNU Radio e da Tecnologia SDR na Aprendizagem das Técnicas de Modulação

Júnio Santos Bulhões[1], Cláudio Afonso Fleury[1], Jeferson Rodrigues Bernardes[1], Ryver Rafael Moreira Franco[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG)
Câmpus Goiânia – GO – Brasil

juniobulhoes@gmail.com, eng.claudio.fleury@gmail.com, jefsrb13@gmail.com,

ryverrafael@gmail.com

Abstract. *The learning of techniques of modulation analogical/digital of signals is so important to training of professionais of telecomunications area. This article evaluated the application of SDR technology using GRC interface GNU Radio platform as auxiliary elements in the teaching of these techniques. Some experiments with modulations and demodulations of signals were performed from using two nodes of connected processing. Initially, the experiments used the wireless sistem that is present in laptops. The signals were transmitted by wireless channel in simplex mode. After that, the same experiment was repeated, however with computers connected through SDR devices (low-cost units) to perform radio transmission. The observed results demonstrated several advantages of using these technologies compared with conventional countertop, they allow a greater range of variation in the parameters of the experiments, increasing versatility, which greatly favors the experimentation in its entirety,besides it promotes a substantial reduction in investments in countertops equipment (economy).*

Resumo. *A aprendizagem de técnicas de modulação analógica/digital de sinais é de grande importância à formação dos profissionais da área de Telecomunicações. Neste artigo avaliou-se a aplicação da tecnologia SDR e da interface GRC da plataforma GNU Radio como elementos auxiliares no ensino das referidas técnicas. Alguns experimentos com modulação e demodulação de sinais foram efetuados a partir do uso de dois nós de processamento conectados. Inicialmente, os experimentos usaram as placas de comunicação sem fio, padrão IEEE 803.11g (unidades originais e embarcadas nos microcomputadores usados). Os sinais foram transmitidos por canal wireless, em modo simplex. Posteriormente, o mesmo experimento foi repetido, porém com os microcomputadores conectados via dispositivos SDR (unidades de baixo custo) para realizar a transmissão sem fio. Os resultados apurados demonstram várias vantagens no uso dessas tecnologias em relação às bancadas convencionais, pois permitem uma maior gama de variação dos parâmetros dos experimentos, aumentando a versatilidade, a qual favorece enormemente a experimentação na sua totalidade, além de promover uma substancial diminuição nos investimentos em equipamentos de bancadas (economia).*

1. Introdução

Com o aumento contínuo do número de usuários de serviços de telecomunicações, vários estudos para tentar melhorar os sistemas já existentes estão em andamento. Uma das tecnologias recentes que demostram maior flexibilidade no desenvolvimento de novas técnicas de telecomunicações é conhecida como Rádio Definido por Software (*Software Defined Radio*,SDR).

Segundo [Rouphael 2009], o SDR é constituído por um equipamento de hardware mínimo e por um software que implementam algumas de suas funções que seriam tradicionalmente implementadas em hardware. O potencial dessa tecnologia propicia a miniaturização e diminuição de custos, além de facilitar o desenvolvimento de projetos de sistemas de telecomunicações [Lima et al. 2016].

A tecnologia SDR é ideal para sistemas que suportam múltiplos protocolos e modulações numa mesma unidade de radiocomunicação, pois podem ser implementados em software, além de ser possível realizar mudanças de versões ou funcionalidades com pequenas intervenções também em software [Reis et al. 2012].

Segundo [Carlos et al. 2016], várias plataformas de software foram desenvolvidas para dar suporte ao desenvolvimento de sistemas de telecomunicações baseados em dispositivos SDR: GNU Radio, OSSIE (*Open Source* SCA (*Service Component Architecture*) *Implementation Embedded*), LabVIEW©(*Laboratory Virtual Engineering Workbench*), SORA (*Software Radio*), IRIS (*Implementing Radio In Software*). Por possuir uma interface gráfica orientada a blocos gráficos que facilita o desenvolvimento de sistemas de telecomunicações, conhecida por GNU Radio Companion (GRC) e por ser gratuito e de código aberto (*open source*), o GNU Radio tem sido escolhido pela maioria dos autores da comunidade científica.

O GRC permite a construção de sistemas via diagrama de blocos, assim como os softwares Simulink©e Labview©, e atualmente possui mais de 150 blocos de processamento[Gandhiraj et al. 2012] que garantem flexibilidade na simulação da maioria dos sistemas de telecomunicações. Cada bloco funcional do GRC é construído por instruções em Phyton que podem ser modificadas a critério do usuário, o qual conta com inúmeras funções prontas disponíveis no pacote GNU Radio.

O ensino de telecomunicações faz uso de equipamentos e instrumentos de alto custo e que em geral não devem ser manipulados diretamente pelos estudantes, gerando barreiras de aprendizagem. O SDR por ser uma tecnologia flexível e de baixo custo, mostra-se bastante viável como alternativa aos equipamentos e instrumentos usuais em bancadas de ensaios. O ensino de técnicas de modulação, filtragem de sinais, análise de espectro, entre outros [Katz and Flynn 2009] podem ser realizados com o apoio de placas SDR.

Este artigo apresenta experimentos de modulações analógicas usando o GNU Radio Companion de forma simples para auxiliar os alunos na aprendizagem de disciplinas em Telecomunicações.

2. Desenvolvimento

2.1. Experimento com Dispositivo SDR

O dispositivo SDR usado nos experimentos como *front-end* de transmissão foi o *HackRF One*, fabricado pela empresa norte americana *Great Scott*[Wooton and Ossmann 2016]. Esse modelo pode operar tanto como transmissor quanto receptor, no modo de comunicação *half-duplex*. A figura 1 apresenta o dispositivo *HackRF One*.

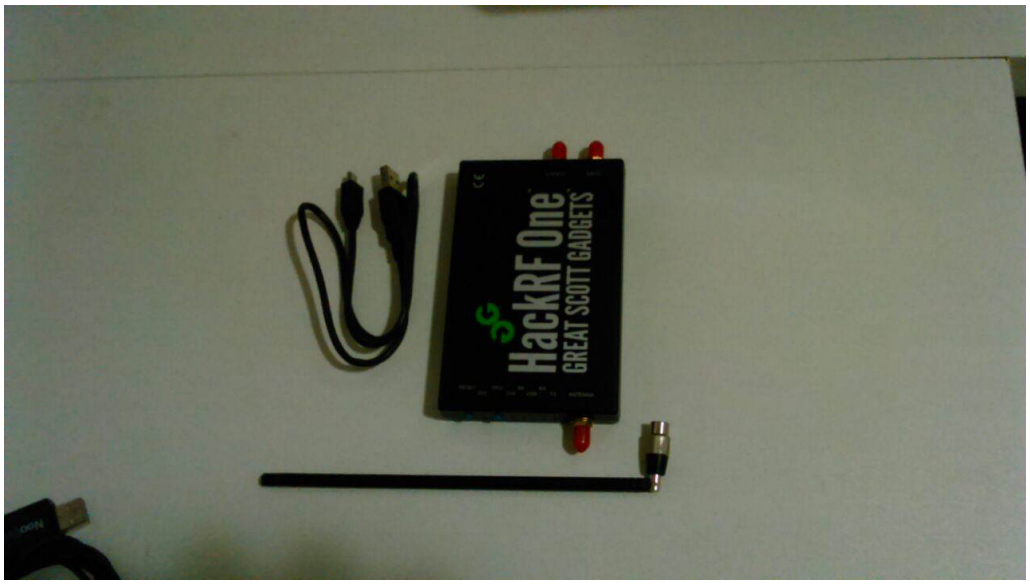


Figura 1. Dispositivo Transceptor *HackRF One*

O dipositivo *HackRF One* foi adotado por possuir operação simplificada (tanto em *driver* de acesso quanto em montagem de HW) e por ser de baixo custo. O modelo usado chega a taxas de amostragem de até 20 milhões de amostras por segundo, além de executar as funções de transmissão ou recepção (não simultâneas). A frequência do oscilador local varia de 10,0 MHz a 6,0 GHz, o que possibilita a realização de experimentos em diversas faixas de frequências, ampliando as possibilidades dos ensaios.

O dispositivo HackRF One tem preço substancialmente menor (aproximadamente U$ 300) quando comparado aos equipamentos e instrumentos de bancada para experimentação em Telecom, porém *front-ends* com mais funcionalidades (dois canais de comunicação e operação como transceptor *full-duplex*), como o conhecido USRP (*Universal Software Radio Peripheral*), podem ser encontrados numa faixa de preços um pouco mais elevada(em torno de U$ 1.500).

Como desvantagens do *front-end* adotado tem-se a limitação do oscilador local que não opera em frequências inferiores a 10 MHz, o que inviabiliza experimentos em frequências baixas, além do fato de não realizar transmissão e recepção simultânea (*full-duplex*).

O dispositivo SDR usado como receptor é denominado Ezcap, e fabricado pela NooElec (empresa com vendas pelo site da Amazom.com). O Ezcap é um dispositivo

com comunicação *simplex* que foi desenvolvido para sintonizar canais da TV digital norte americana (padrão DVB).A figura 2 apresenta o dispositivo Ezcap.

O tamanho reduzido do Ezcap é um dos detalhes que chama a atenção nesse dispositivo, sua aparência é similar a de um pendrive com um conector USB (alimentação e saída do sinal em frequência intermediária) e outro conector para antena de captação do sinal de RF. O papel desempenhado por este dispositivo é transladar sinais com frequências próximas à da portadora sintonizada para uma frequência menor (frequência intermediária), e depois o sinal transladado é filtrado, passando as frequências baixas e por último é realizada a digitalização do sinal analógico para ser entregue ao dispositivo de processamento digital(microcomputador).

Esse receptor tem custo bem inferior ao do HackRF One (aproximadamente U$ 17). Existem vários dispositivos SDR similares ao Ezcap, baseados no circuito integrado RTL2832U (fornece sinais amostrados em quadratura, resolução de 8 bits). Embora este dispositivo não seja fabricado no Brasil o mesmo pode ser adquirido em alguns websites locais.

Figura 2. Dispositivo Receptor Ezcap

Alguns pontos fracos do Ezcap: gera no máximo 3,2 milhões de amostras por segundos, o que inviabiliza trabalhar com sinais de altas resoluções; a faixa de frequência

de recepção varia de 24 MHz a 1,7 GHz. A frequência mínima de operação de 24 MHz inviabiliza estudos com ondas curtas, e a frequência máxima de operação de 1,7 GHz não permite trabalhar com a tecnologia Bluetooth que opera normalmente na faixa ISM de 2,4 GHz. Pelo fato de operar apenas como receptor este dispositivo inviabiliza comunicações bilaterais (full-duplex).

2.2. GNU Radio e SDR

O software GNU Radio foi desenvolvido para operar em sistemas operacionais Linux, embora já existam versões para outras plataformas, tais como Windows e OS-X.

A plataforma Linux indicada para estudos iniciais é a plataforma Ubuntu, por ela possuir uma grande comunidade na Internet que atua como suporte técnico em eventuais erros. O GNU Radio não é encontrado no Ubuntu Software e por isso sua instalação se torna um pouco mais complexa, necessitando de vários comandos no terminal do sistema operacional Ubuntu.

Para se usar dispositivos SDR no GNU Radio é necessário a instalação de algumas bibliotecas para fazer a conexão desses dispositivos com o SW. Além disso, existe pouca ajuda do GRC a respeito dos parâmetros encontrados nos blocos de funções, dificultando a configuração adequada dos mesmos.

Visando facilitar o acesso ao GNU Radio e suas funcionalidades o grupo Corgan Labs desenvolveu o projeto GNU Radio Live Environment, o qual disponibiliza todas as ferramentas citadas no parágrafo anterior, e de uma forma bem mais simples [Corgan 2016].

O *GNU Radio Live SDR Environment* foi criado para a distribuição Linux Ubuntu que contém todas as ferramentas necessárias para se trabalhar com SDR e GNU Radio. O programa GNU Radio Companion, os blocos funcionais e os drives para acesso aos dispositivos SDR são nativos nesse projeto.

O GNU Radio Companion é uma ferramenta gráfica que se utiliza de blocos funcionais para executar diversos processamentos de sinais. O fluxo de informações gerado pelas conexões dos blocos facilita o entendimento do processamento dos sinais no todo, assim como o papel desempenhado por cada bloco. Além disso, o GRC possui vários exemplos de sistemas usuais em telecomunicações que podem auxiliar os usuários no desenvolvimento de seus próprios sistemas em diagrama de blocos.

A figura 3 apresenta o software GNU Radio Companion. O botão onde o retângulo número 1 se encontra é para criar um novo fluxograma. O retângulo número 2 é para abrir um fluxograma já existente e o retângulo número 3 é para salvar o fluxograma construído. O local para se contruir os fluxogramas está representado pelo retângulo 8. Os blocos existentes no GRC podem ser encontrados na lateral direita(retângulo 9). O ícone onde se situa o retângulo 4 encontra-se em vermelho caso haja alguma irregularidade no fluxograma.

Com o fluxograma criado o mesmo deve ser gerado e executado, esses processos são realizados pelos botões situados nos retângulos 5 e 6 respectivamente. Algumas informações do atual estado do GRC podem ser encontrados na parte inferior do programa, e por fim para encerrar a execução do fluxograma, o botão localizado no retângulo 7 deve ser clicado.

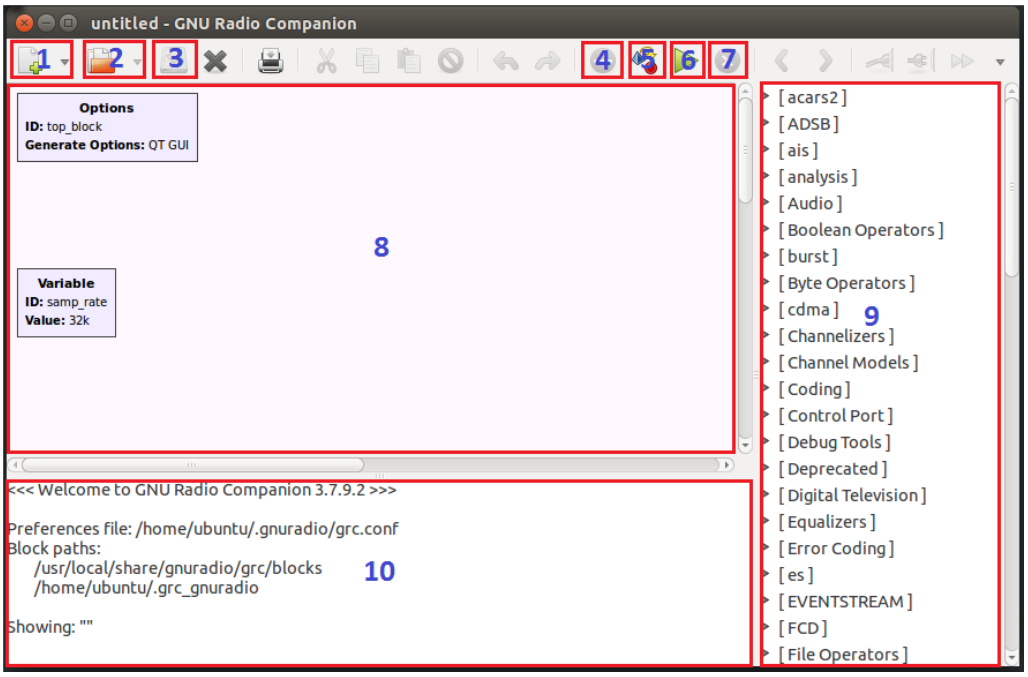


Figura 3. GNU Radio Companion

Depois de desenhado o diagrama de blocos correspondente ao sistema de interesse, ele é implementado com a geração de um *script* em Ling. Python, e consequente simulação (execução do *script*). A etapa de conversão usa como entrada o diagrama de blocos com todas as suas conexões para gerar o *script* que representa todos os elementos gráficos (blocos funcionais). O usuário pode editar o arquivo do *script* para adequá-lo aos requisitos do projeto ou otimizá-lo de tal modo que a conversa automática não consiga por si, propiciando versatilidade aos sistemas convertidos. A execução do programa correspondente ao sistema projetado pode ser feita na própria interface do GRC ou em qualquer computador que possua GNU Radio instalado através do terminal.

Durante os estudos relativos ao GRC deparou-se com algumas dificuldades para conseguir materiais consistentes sobre a utilização do GRC, e mais especificamente sobre o uso básico e avançado de cada bloco, assim como informações mais detalhadas sobre cada parâmetro disponível em cada bloco funcional.

2.3. Experimentos de Telecom

A avaliação da tecnologia SDR e do software GNU Radio Companion passou pela realização de alguns experimentos de validação. Os ensaios realizados trabalharam com várias técnicas de modulação e também pela verificação da taxa de erro de bits (*Bit Error Rate* - BER) para as modulações digitais.

As técnicas de modulação analógica utilizadas foram a modulação em amplitude (*Amplitude Modulation* - AM) e a modulação em frequência (*Frequency modulation* - FM), enquanto que as modulações digitais foram a modulação por deslocamento de frequência gaussiana (*Gaussian Frequency Shift Keying* - GFSK), a modulação por deslocamento mínimo gaussiano (*Gaussian Minimum Shift Keying* - GMSK), a modulação por deslocamento de fase (*Phase Shift Keying* - PSK) e a modulação de amplitude em

quadratura (*Quadrature Amplitude Modulation* - QAM).

A análise da BER foi estudada para as modulações PSK e QAM. Os ensaios para essas técnicas tiveram pequenas alterações nos *scripts* Python gerados pelo GRC, para que os mesmos fossem capazes de modificar os parâmetros de forma automática e salvar os resultados encontrados. Essa facilidade demostra a flexibilidade da ferramenta GNU Radio e de sua interface gráfica, o GRC.

2.4. Metodologia Aplicada

A análise da tecnologia SDR com a utilização do software GNU Radio para simular as técnicas de modulação se deu a partir da realização de quatro ensaios diferentes, com o objetivo de validar o uso dessas ferramentas no ensino-aprendizagem de telecomunicações.

O primeiro ensaio fez a transmissão de informações virtualmente dentro do software GNU Radio. Este caso particular envolve todas as etapas pertinentes à modulação e à demodulação, porém utilizando-se canal ideal (sem ruído e sem distorções).

A segunda parte dos ensaios fez uso das placas de rede Ethernet sem fio dos dois microcomputadores para realizar a transmissão e a recepção de sinais. Esta segunda situação utiliza-se de um canal de dados onde já existe uma modulação digital para transmissão de bits, ou seja, o sinal modulado foi remodulado para ser transmitido ao outro nó computacional.

A terceira etapa consistiu na substituição do sistema de rede *Ethernet wireless* por dispositivos SDRs que fizeram as etapas de transmissão do sinal após o sinal ser modulado e da recepção e posterior demodulação. Esta etapa demonstrou o uso do sistema de transmissão num canal real (ruídos e distorções). A transmissão foi realizada pelo dispositivo SDR *HackRF One* e a recepção, pelo dispositivo SDR Ezcap.

O quarto experimento simulou canais com ruídos variados e sinal de energia constante para se analisar a taxa de erro de bits (BER) proporcionada por cada uma das modulações digitais pretendidas.

3. Resultados

3.1. Primeiro Tipo de Experimento

Os diagramas de blocos para simular uma comunicação em um único nó computacional foram de fácil implementação. Para algumas modulações testadas como GMSK, QPSK, o GRC já possuía exemplos de diagramas implementados onde a ideia principal pôde ser replicada para a construção dos diagramas de blocos destinados às demais modulações digitais que não eram contempladas com exemplos.

A análise dos resultados contou com a ajuda de blocos que simulam aparelho de medição com características similares a de um osciloscópio. Através desses blocos, observam-se as características do sinal, como amplitude, frequência, fase e potência, as quais são mostradas em gráficos com taxa de atualização controlada pelo usuário (microssegundos). O GRC conta com duas famílias distintas de blocos para desenhos gráficos que derivam das bibliotecas para interface gráfica em Python: QTPython e WXPython.

Analisando o desempenho dos blocos que implementam o QTPython e o WXPython percebe-se que o QTPython possui um tratamento prévio do sinal para melhor

visualização das informações disponíveis, contudo seu plano de fundo é escuro (embora possa ser alterado, isso exigirá conhecimentos maiores sobre a biblioteca em uso) o que não é aconselhado para trabalhos acadêmicos como relatórios e artigos. O WXPython por sua vez, apresenta interface gráfica limpa (cor de fundo branca, também pode ser alterada) que é mais indicada aos trabalhos científicos.

A figura 4 apresenta o sinal modulado em frequência recebido virtualmente. Percebe-se que o sinal encontra-se com amplitude constante e frequência variavel.A presença de ruído não foi observada,já que a transmissão e recepção virtual utilizam-se de meio físico ideal (sem ruído).

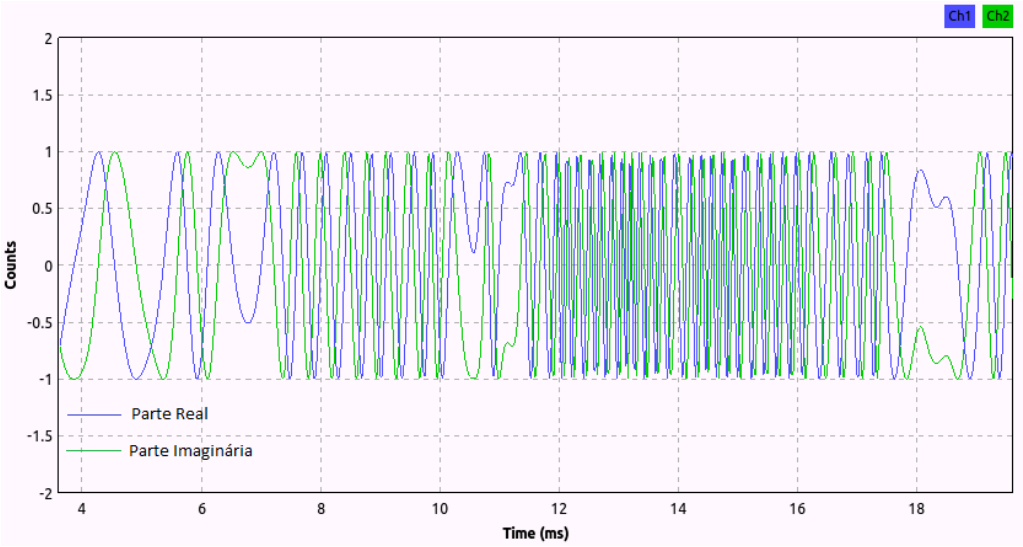


Figura 4. Sinal Modulado em Frequência

3.2. Segundo Tipo de Experimento

Com os modelos de diagramas de blocos de modulação e demodulação criados para a realização de comunicação virtual já implementados e avaliados nos experimentos anteriores, o desafio do segundo tipo de experimento estava presente em como realizar comunicação entre os dois computadores utilizando-se da conexão *wireless*.

As dificuldades do segundo tipo de experimento vêm do fato de que não se encontrou exemplos implementados no GRC, assim como a falta de literatura sobre este tipo de ensaio. Embora o GRC não possua exemplos adequados disponíveis, os blocos do tipo TCP/IP e UDP são nativos no GRC, o que evitou a necessidade da implementação de blocos para realizar esse tipo de processo, pelo menos numa etapa inicial de aprendizagem do GRC.

A figura 5 apresenta o sinal modulado em amplitude(AM) recebido através do sistema *wireless*. Como na transmissão virtual, os sinais recebidos utilizando as placas de rede *wireless* não apresentaram ruído.

Para modulações analógicas os resultados alcançados demostram similaridade aos realizados virtualmente, porém para modulações digitais a quantidade de informação a ser transmitida inviabilizou sua análise para o arquivo de áudio adotado. O problema não

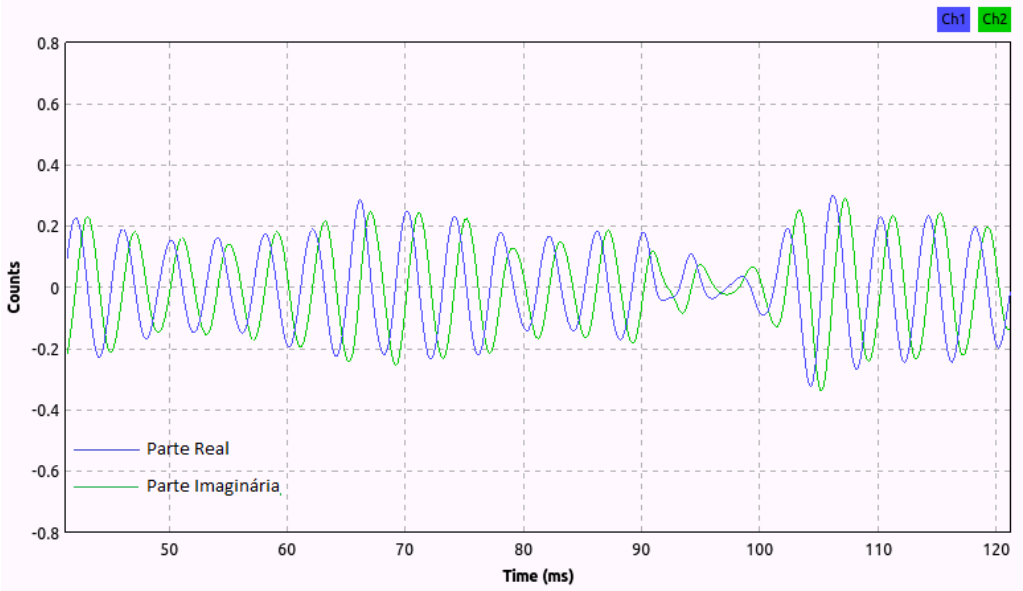


Figura 5. Sinal Modulado em Amplitude Captado pelo Computador Receptor

era esperado, porém após uma análise dos sintomas, constatou-se que o sistema wireless não suportava o tráfego de dados necessário nas modulações digitais. Para transmitir sinais com modulação digital o arquivo de áudio foi substituído por arquivos de texto e fotos que não ultrapassavam 2 MB. A transmissão e recepção digital dos arquivos tiveram taxa de transferência reduzida para se evitar os problemas citados acima e os resultados alcançados foram compativeis com os obtidos nas transmissões analógicas.

3.3. Terceiro Tipo de Experimento

O terceiro tipo de experimento consiste na utilização dos dispositivos SDR para realizarem a comunicação entre o nó transmissor e o nó receptor, tendo sido desligadas as placas *wireless* de cada microcomputador. A figura 6 apresenta os dispositivos SDR e computadores envolvidos na realização desses experimentos.

Figura 6. Sistema de Transmissão Utilizando Dispositivos SDR

As modulações digitais QAM e PSK não geraram resultados satisfatórios. Detectou-se que o problema foi a baixa taxa de amostragem conseguida pelo dispositivo receptor (Ezcap) e que impedia a demodulação adequada da informação a partir do sinal captado pelo dispositivo receptor. Para as demais modulações, observou-se a presença de ruído no canal, fenômeno que já era esperado. A presença de ruído impediu a demodulação do sinal em alguns instantes.

A figura 7 apresenta o sinal modulado em GMSK recebido pelo *dongle* Ezcap. Percebe-se que o mesmo contém ruido, pois existe uma irregularidade que pode ser observada na forma de onda analisada. Isto já era esperado, pois como o canal utilizado(ar) é repleto de fenômenos que interferem no sinal original gerando o ruído.

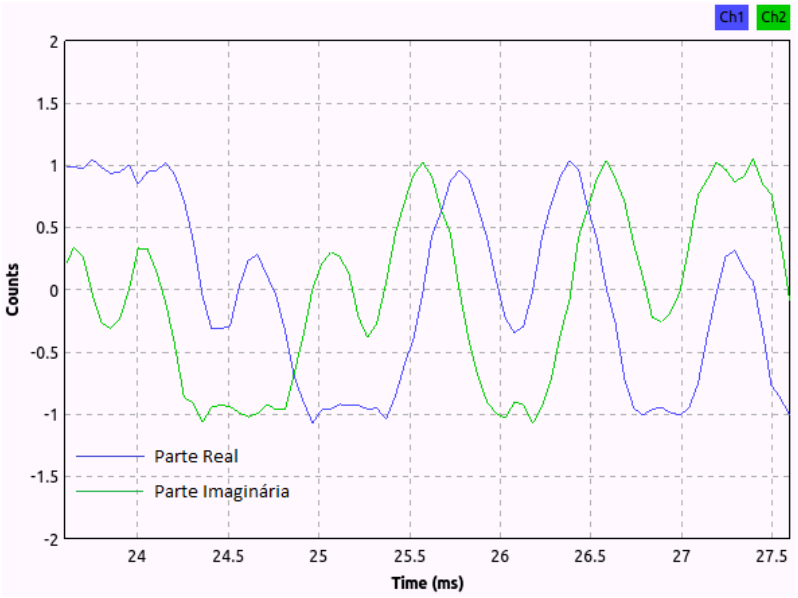


Figura 7. Sinal Modulado em GMSK Captado pelo Ezcap

Outro experimento utilizando os dispositivos SDR foi a sintonia de uma estação de rádio comercial FM local e a inserção de um sinal modulado em frequência na mesma frequência sintonizada pelo receptor. O sistema receptor mudou o sinal de informação coletado automaticamente, pois o sinal transmitido pelo dispositivo transmissor (*HackRF One*) tinha potência maior no local onde encontrava-se o dispositivo receptor (Ezcap) do que a estação de rádio comercial sintonizada.

3.4. Quarto Tipo de Experimento

O quarto tipo de experimento propôs uma análise da BER para as modulações QAM e PSK. Neste tipo de experimento a modificação do arquivo Python gerado pelo GRC foi necessária devido a alta taxa de dados que se pretendia utilizar. A figura 8 apresenta as curvas da BER levantadas através da simulação para as modulações QAM e PSK.

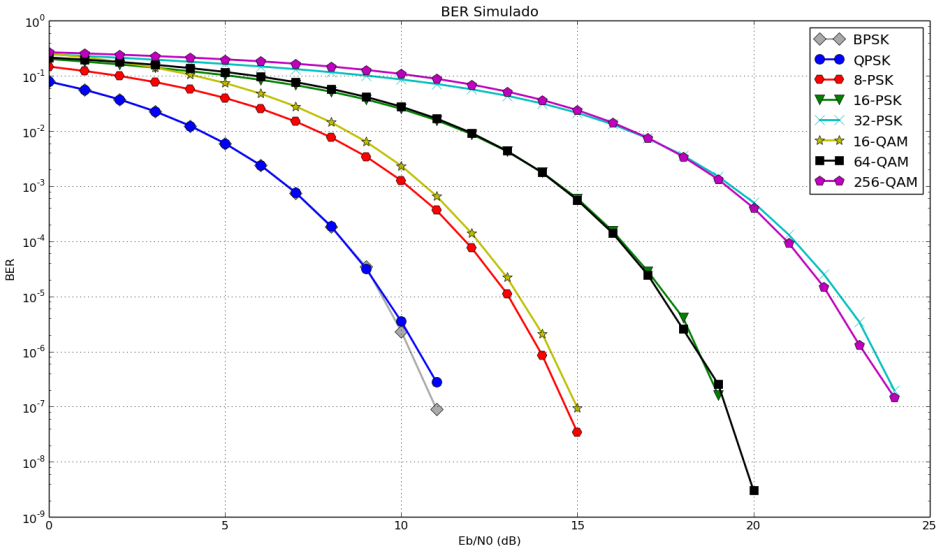


Figura 8. Curvas da BER Simuladas

A simulação realizada provou que a modulação 16-QAM e 8-PSK possuem curvas BER similares. Isto demonstra que a modulação 16-QAM é mais eficiente do que a 8-PSK, pois a mesma consegue transportar mais bits por unidade de tempo e consequentemente reduzir o tempo necessário da transmissão.

Uma generalização pode ser feita ao se analisar as curvas BER para as modulações 8-PSK, 16-PSK, 32-PSK, 16-QAM, 64-QAM e 256-QAM. A modulação QAM usa mais eficientemente o canal do que a PSK. Isto indica que a modulação QAM pode substituir sistemas que utilizam PSK com um ganho significativo.

Não houve dificuldades para se modificar os arquivos Python gerados pelo GRC, uma vez que existe uma grande comunidade de programadores Python na Internet, e que é rica na solução de problemas variados.

4. Conclusão

O GNU Radio demonstrou ser uma ferramenta de interface bastante amigável e de fácil entendimento, tornando simples a montagem de sistemas de modulações AM e FM, com poucos blocos necessários à montagem dos diagramas de blocos para o transmissor e receptor. Por não possuir um bloco de modulação em amplitude disponível em sua biblioteca, foi necessário o desenvolvimento de uma estrutura com os blocos existentes para se construir um modulador em amplitude.

A transmissão e a recepção virtual possuem custo de montagem zero (nenhum HW externo envolvido), podendo ser realizado em qualquer local que possua um computador com o GNU Radio Companion instalado. Isto é muito útil na simulação das diversas etapas de cada modulação e demodulação e na análise dos resultados.

A transmissão e a recepção sem fio apresentaram resultados similares aos obtidos com o primeiro experimento (transmissão e recepção virtuais), porém nesta etapa tem-se uma atenção maior na simulação da comunicação por se tratar de dois nós computacionais distintos, facilitando a aprendizagem dos fenômenos envolvidos no processo por parte dos estudantes.

A transmissão e a recepção utilizando dispositivos SDR demonstraram como a tecnologia SDR funciona. O dispositivo transmissor estava hora transmitindo informações moduladas em amplitude, hora em frequência, sem que fosse necessário trocar os dispositivos *front-end*. Este tipo de ensaio pode ser utilizado em laboratórios para demonstrar transmissões e recepções reais de dados ou informações analógicas (fala, áudio, vídeo). Outra vantagem é que se pode testar diferentes modulações, filtros e parâmetros, sem que seja necessário adquirir outros equipamentos (flexibilidade da tecnologia SDR) ou refazer novo circuito na bancada, bastando apenas a troca do arquivo do GRC (diagrama de blocos).

Pretende-se no futuro próximo realizar a confecção de diagramas de blocos no GRC para implementar modulações digitais para facilitar o entendimento dessas de modo mais natural e flexível.

Referências

Carlos, E. H., Munhóz, F. S., and Naressi, L. C. (2016). Um estudo sobre rádio definido por software (rds). *Instituto Federal de Educação, Ciência e Tecnologia de Goiás*.

Corgan, J. (2016). Using the gnu radio live sdr environment. *Available from World Wide Web: http://gnuradio.org/blog/using-gnu-radio-live-sdr-environment/*.

Gandhiraj, R., Ram, R., and Soman, K. (2012). Analog and digital modulation toolkit for software defined radio. *Procedia Engineering*, 30:1155–1162.

Katz, S. and Flynn, J. (2009). Using software defined radio (sdr) to demonstrate concepts in communications and signal processing courses. In *2009 39th IEEE Frontiers in Education Conference*, pages 1–6. IEEE.

Lima, G. C., Carvalho, S. P., and Oliveira, S. J. S. (2016). Estudo de desenvolvimento de sistemas de rádio definido por software. *Pontifícia Universidade Católica de Goiás*.

Reis, A. L. G., Barros, A. F., Lenzi, K. G., Meloni, L. G. P., and Barbin, S. E. (2012). Introduction to the software-defined radio approach. *IEEE Latin America Transactions*, 10(1):1156–1161.

Rouphael, T. J. (2009). *RF and digital signal processing for software-defined radio: a multi-standard multi-mode approach*. Newnes.

Wooton, E. and Ossmann, M. (2016). Great scptt gadgets. *Available from World Wide Web: https://greatscottgadgets.com/*.

A Olimpíada Brasileira de Informática como forma de atrair estudantes do Ensino Médio aos cursos de Tecnologia da Informação

Adriano H. Braga[1], Ramayane B. Braga[1], Gustavo S. Faquim[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano)
Caixa Postal 51 – 76.00-000 – Ceres – GO – Brasil

{adriano.braga,ramayane.santos}@ifgoiano.edu.br,
gustavofaquim408@gmail.com

Abstract. *Computation has been increasingly being inserted in daily activities, but many people have knowledge only about their use and not operation. Programming language is one of the disciplines aimed at this study and is seen by many students of the area as one of the hardest to learn and consequently causing high school dropout. With the inclusion of this discipline in high school courses, through especially by the Federal Institutes and the application of OBI, this paper aims to demonstrate this issue as a way to captivate the interest of students to the eld of Information Technology.*

Resumo. *A computação tem sido cada vez mais sendo inserida nas atividades diárias, porém muitos possuem conhecimento apenas quanto a sua utilização e não quanto ao seu funcionamento. Linguagem de programação é uma das disciplinas que visa este estudo e é tida por muitos estudantes da área como uma das mais difíceis de aprendizado e consequentemente que causam maior evasão escolar. Com a inserção desta disciplina nos cursos de Ensino Médio, por meio principalmente dos Institutos Federais e com a aplicação da OBI, este artigo visa demonstrar esta problemática como uma forma de cativar o interesse de estudantes para a área de Tecnologia da Informação.*

1. Introdução

Anualmente, a Sociedade Brasileira de Computação (SBC) organiza duas competições nacionais: a Olimpíada Brasileira de Informática (OBI), com foco nos alunos do ensino fundamental e do ensino médio, e a Maratona de Programação com foco nos alunos do ensino superior. A OBI é uma competição organizada de forma similar as demais olimpíadas cientí cas brasileiras, como a Olimpíada Brasileira de Matemática de Escolas Públicas (OBMEP) e a Olimpíada Brasileira de Astronomia e Astronáutica (OBA). A organização da OBI, em âmbito nacional, está a cargo do Instituto de Computação da UNICAMP [OBI 2016].

A OBI é dividida em três modalidades, sendo:

- **Modalidade Iniciação:** são aplicadas questões de múltipla escolha, cada uma com cinco alternativas e apenas uma correta, abordando problemas de lógica a serem solucionados sem a utilização de um computador. É destinada a estudantes regularmente matriculados no Ensino Fundamental;
- **Modalidade Programação:** com a utilização de computador, destinada a estudantes regularmente matriculados no Ensino Fundamental e Médio, em que os competidores devem solucionar problemas computacionais que envolvem conteúdos como: estruturas de dados e técnicas de programação;
- **Modalidade Universitária:** ocorre de forma similar a modalidade Programação, contudo é destinada apenas a estudantes que estejam regularmente matriculados pela primeira vez no primeiro ano de um curso de graduação.

Para todas as etapas e modalidades, os competidores devem realizar a prova individualmente. Nas modalidades Programação e Universitária os competidores podem solucionar o algoritmo proposto utilizando as seguintes linguagens de programação: Pascal, C, C++, Python, Java ou Javascript. A prova deve ser realizada em computador com ambiente de programação adequado sem conexão com a Internet e sem consulta a quaisquer material de consulta.

A modalidade de programação possui três níveis, sendo um nível especí co para os alunos do Ensino Fundamental, nível júnior; outro nível para os alunos do Ensino Fundamental e alunos até a 2ª série do Ensino Médio, nível 1; e um nível para alunos do Ensino Fundamental ou Ensino Médio, nível 2. Para esta modalidade é restrito a inscrição apenas a alunos que não tenham completado 20 anos até o primeiro dia de julho do ano de sua inscrição na olimpíada. Neste artigo são abordados resultados de estudantes do Campus Ceres do IF Goiano que realizaram os Níveis 1 e 2 nos anos de 2015 e 2016.

A OBI é realizada em duas fases, sendo que na primeira fase todos os competidores inscritos do Brasil competem simultaneamente em prova de mesmo nível, dia e horário conforme de nido previamente pela organização nacional. Os melhores classi cados em cada nível da Fase 1, são convocados para a próxima fase, perfazendo uma média geral de dez por cento do quantitativo total de competidores de todas as escolas que realizaram a prova.

Os melhores classi cados após a realização da Fase 2, recebem medalhas nas categorias ouro, prata e bronze, além de honra ao mérito. Dos competidores da Modalidade Programação Nível 2 são convidados os primeiros colocados com maior pontuação nal das duas fases para participarem de um treinamento na UNICAMP com vistas em se prepararem para a Olimpíada Internacional de Informática (IOI, do inglês – *International Olympiad in Informatics*). Após esta capacitação e a realização de uma

nova seleção conforme o rendimento dos alunos, são de nidos os quatro alunos que representarão o Brasil na edição internacional.

A IOI é atualmente a segunda maior olimpíada cientí ca do mundo, cando atrás apenas da Olimpíada Internacional de Matemática (IMO, do inglês *International Mathematical Olympiad*), e obteve na última edição em agosto deste ano de 2016 na Rússia, a participação de 80 países e mais de 300 competidores. A IOI possui como um de seus patronos a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO).

Um dos principais objetivos da OBI é despertar nos alunos o interesse pela ciência da computação na formação básica e consequentemente diminuir a crescente falta de pro ssionais na área de Tecnologia da informação que vem sendo repercutida a nível mundial. Segundo dados publicados pela Brasscom, Associação Brasileira das Empresas de Tecnologia da Informação e Comunicação, em 2020 o Brasil passará por uma carência de mais de 750 mil pro ssionais da área [Brasscom 2016]. As causas para a carência desses pro ssionais vão desde o baixo interesse dos estudantes brasileiros por ciências exatas até o alto índice de evasão dos alunos nos cursos ligados à tecnologia.

Segundo a Catho [Catho 2015], mesmo com o país em crise, ocorreu o aumento em mais de 44% no número de vagas no setor de tecnologia durante o primeiro semestre do ano de 2015 se comparado ao ano de 2014. Conforme Daniel Rodrigues Costa, presidente da empresa Catho: “As empresas precisam reduzir custo, melhorar processo, gerar novas oportunidades de negócios e tecnologia hoje resolve isso” [Catho 2015].

2. O ensino de linguagem de programação

Os componentes curriculares que envolvem a subárea de conhecimento Linguagens de Programação, da Ciência de Computação, está presente na maioria dos cursos relacionados a Tecnologia da Informação, sejam estes cursos de nível médio ou superior.

O aprendizado de linguagem de programação é de extrema importância aos estudantes de Informática, sendo que tal conhecimento está relacionado de forma direta ou indireta com todas as demais subáreas, contudo é ainda responsável pelas disciplinas de maior reprovação dos cursos de Informática e que exigem uma abstração de difícil assimilação por práticas diárias do mundo real [Sudol 2011].

Linguagem de programação é destacada por ser um dos principais motivos em aumentar da evasão dos estudantes dos cursos de Tecnologia da Informação, possuindo um dos maiores índices de reprovação entre as demais subáreas [Silva 2011]. E além disso, nem sempre é reconhecida como de fato sua própria descrição induz, uma “linguagem”, assim como a língua portuguesa, inglesa, de sinais e outras. Tal componente curricular é responsável por responder aos estudiosos como os computadores se codi cam/comunicam. Desta forma, com o intuito de promover a

realização da OBI e como consequência o ensino de linguagem de programação no Campus Ceres do IF Goiano, foi aprovado um projeto de pesquisa via Edital CNPq-SETEC/MEC no nal do ano de 2014 que vem sendo executado desde então.

A disciplina de Linguagem de Programação está presente em todos os cursos de Informática do Campus Ceres, como também na maioria dos cursos da área de nível técnico integrado ao ensino médio presentes no país por meio de em sua maioria dos Institutos Federais.

3. Os Insitutos Federais

A Rede Federal de Educação Pro ssional, Cientí ca e Tecnológica tem como data de seu surgimento o ano de 1909, quando o então Presidente da República, Nilo Peçanha, instituiu por meio de um decreto presidencial a criação de 19 escolas de Aprendizes e Artí ces, tida como instrumento de política no momento voltado para as "classes desprovidas" [Pacheco 2010; Rede Federal 2016].

Com o passar dos anos, mais precisamente em 1978, três escolas federais, e posteriormente na década de 90 várias outras escolas técnicas e agrotécnicas federais foram transformadas em Centros Federais de Educação Tecnológica (CEFET), formando assim a base desta atual Rede Federal de Educação Pro ssional, Cientí ca e Tecnológica cobrindo todo o território nacional e prestando um serviço a nação como importante estrutura para que todas as pessoas tenham efetivo acesso às conquistas cientí cas e tecnológicas e possam se quali car pro ssionalmente para os diversos setores da economia brasileira [Pacheco e Rezende 2009].

Em 2008, a Escola Agrotécnica Federal de Ceres foi transformada em Campus do Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano), em função da reestruturação da rede federal de educação pro ssional e tecnológica, que destituiu 31 centros federais de educação tecnológica (Cefets), 75 unidades descentralizadas de ensino (Uneds), 39 escolas agrotécnicas, 7 escolas técnicas federais e 8 escolas vinculadas a universidades por meio da Lei 11.892 de 29 de dezembro de 2008 para então a consolidação e expansão da atual Rede Federal de Educação Pro ssional, Cientí ca e Tecnológica, conforme Figura 1.

Após essa reorganização da rede federal de educação, o Campus Ceres do IF Goiano passou a ter o compromisso de gerar, promover e difundir conhecimentos cientí cos e tecnológicos, para o desenvolvimento sustentável das comunidades e dos segmentos socioeconômicos de sua área de abrangência. Para isso, tem ofertado cursos de formação pro ssional, em nível de ensino médio integrado ao ensino técnico, até o nível superior com cursos tecnológicos, cursos de bacharelado e cursos de pós-graduação Lato Sensu (especialização e aperfeiçoamento) e Stricto Sensu (mestrado e doutorado).

Figura 1. Mapa da Rede Federal de Educação Pro ssional, Cientí ca e Tecnológica [Instituições da Rede 2016]

3.1. A OBI no Campus Ceres do IF Goiano

A cidade de Ceres é considerada uma cidade polo no Vale do São Patrício no que se refere à prestação de serviços de educação e saúde, contudo até o ano de 2015 o Campus Ceres do IF Goiano não havia sediado nenhum em evento em Tecnologia da Informação e a região não havia representantes em competições cientí cas da área.

Neste sentido, este trabalho incita aos alunos exercitarem o conhecimento estimulando assim o interesse pela Computação e Ciência em geral, além de fomentar a articulação entre o projeto pedagógico dos cursos alinhado à prática de programação requisitada por empresas e demanda de serviços de Tecnologia da Informação do Vale do São Patrício, ressaltando que cerca de 25% (335) de todos os estudantes desta unidade estão atualmente matriculados em cursos da área de Informática.

Em 2015 principalmente com o intuito de inserir a cidade no cenário nacional, foi realizado pela primeira vez a OBI na cidade de Ceres, organizada localmente pelos professores do Campus Ceres do IF Goiano. Além da aplicação da prova, também foi disposto um formulário prévio online para treinamento dos que se inscreveram no mesmo. Ao nal foram um total de 80 estudantes atendidos, sendo que todos estes estavam regularmente matriculados no curso Técnico em Informática Integrado ao Ensino Médio da própria unidade.

Do total dos 80 estudantes inscritos:

- 40 estudantes estavam regularmente matriculados na 1ª série do Ensino Médio;
- 20 estudantes estavam regularmente matriculados na 2ª série do Ensino Médio;
- e, 20 estudantes estavam regularmente matriculados na 3º série do Ensino

Médio.

As aulas foram realizadas durante os meses de Maio e Junho nos períodos vespertinos em que os estudantes não possuíam aulas regulares, sendo que cada aula possuía 2 horas de duração e era destinada aos estudantes respectivamente matriculados na mesma série. Com o intuito de que não atrapalhasse nas atividades escolares os agendamentos eram remarcados quando possível ciente de que os interessados estudam em tempo integral na escola, possuem cerca de 16 disciplinas por série e a maioria não moravam na cidade de Ceres.

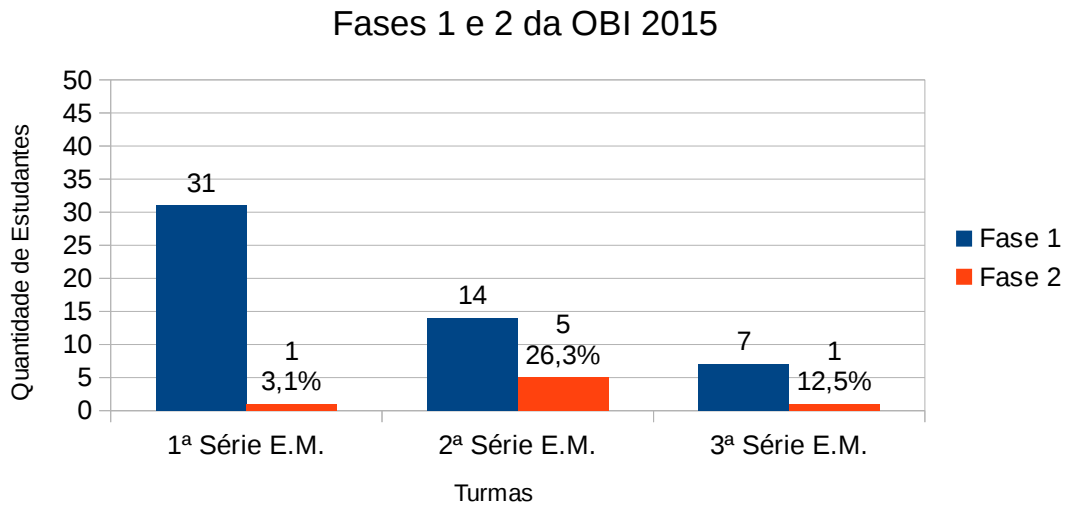
Tendo que as aulas eram destinadas para cada grupos de alunos matriculados respectivamente na mesma série, e como combinado previamente com o professor da disciplina de Lógica de Programação, a qual está presente na matriz curricular da 1ª série do Ensino do Médio Integrado, foi de nido portanto a linguagem de programação Python para que os alunos pudessem aproveitar os conhecimentos adquiridos durante a disciplina no treinamento da OBI.

Já para os estudantes matriculados na 2ª e 3ª série do Ensino Médio, como haviam aprendido Portugol na disciplina de Lógica de Programação e após então a linguagem de programação PHP nos anos seguintes, não foi possível aproveitá-las por não serem linguagens aceitas pela competição. Em discussão na aula inicial com os interessados e elencando as linguagens de programação aceitas pela OBI, foi de nido a linguagem de programação C para o treinamento. Vale ressaltar que embora estes estudantes possuíam conhecimento de programação, não estavam habituados com a resolução de algoritmos computacionais como os propostos pela OBI.

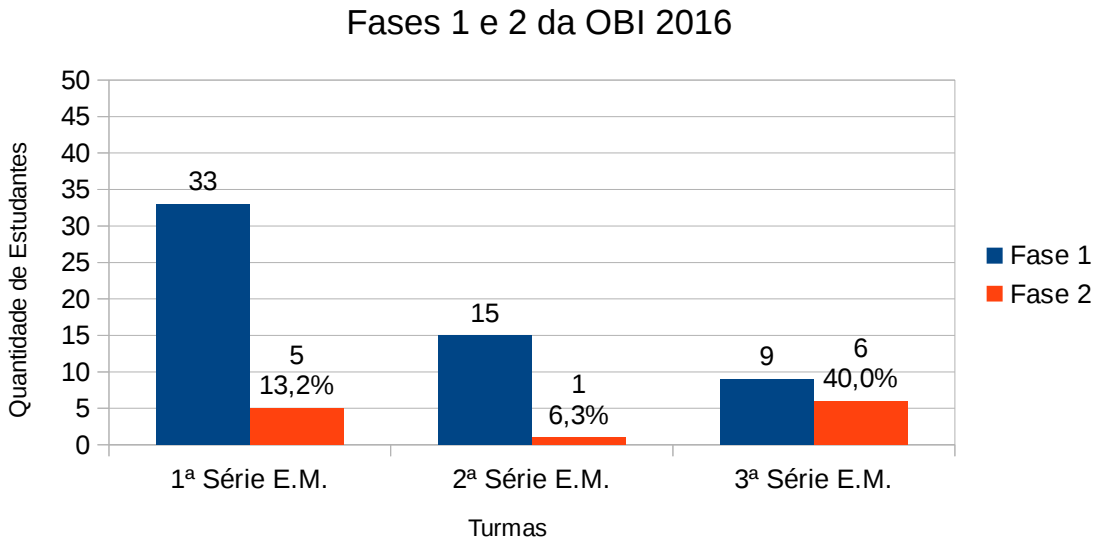
Com o passar do treinamento e devido à proximidade do período de aplicação das veri cações de aprendizagem, como também falta de equipe para o treinamento e outros motivos ainda não avaliados, a quantidade de alunos presentes nos treinamentos foram diminuindo, tendo que ao nal menos de 30% do público inscrito estava presente nas aulas. Na Fase 1 da OBI 2015, um total de 52 estudantes do Campus Ceres do IF Goiano realizaram a prova, sendo que 45 realizaram o Nível 1 (1ª e 2ª Série) e 7 o Nível 2 (3ª Série), conforme o Grá co 1.

Os resultados da Fase 1 foram bastante satisfatórios pois a Escola se gurou como a com o maior número de aprovados na modalidade programação para a Fase 2 da região Centro-Oeste, foram 13% (07) aprovados dos estudantes que zeram a prova.

Em 2016 houve um acréscimo de 9% no número de inscritos para o treinamento, contudo o número de participantes efetivos permaneceram diminuindo durante o curso. Neste ano a divulgação e os treinamentos se iniciaram no mês de Março e a linguagem adotada para todas as turmas foi o Python, ressalvando um tempo extra para plantão de dúvidas em outras linguagens de programação. Na Fase 1 da OBI 2016, um total de 57 estudantes do Campus Ceres do IF Goiano realizaram a prova, sendo que 33 realizaram o Nível 1 (1ª Série) e 24 o Nível 2 (2ª e 3ª Série), conforme o Grá co 2.



Grá co 1. Quantidade de estudantes do IF Goiano Campus Ceres que realizaram a Fase 1 e se classi caram para a Fase 2 na OBI de 2015



Grá co 2. Quantidade de estudantes do IF Goiano Campus Ceres que realizaram a Fase 1 e se classi caram para a Fase 2 na OBI de 2016

Mais uma vez o resultado foi satisfatório pois a Escola obteve o maior número de aprovados na modalidade programação para a Fase 1 da região Centro-Oeste, foram 21% (12) aprovados dos que realizaram a prova.

4. Metodologia

Para que o treinamento e a aplicação das provas da OBI fossem realizadas, foram necessários 2 laboratórios de Informática da unidade que possuem sistema operacional livre, Ubuntu. Não se fez necessário a solicitação de nenhum recurso extra, pois todos

os softwares utilizados eram livres e já estavam instalados nos computadores. Participaram durante os dois anos de projeto: 3 professores e 65 estudantes, sendo que 4 destes estudantes atuaram como monitores no ano de 2016.

As aulas foram ministradas presencialmente nos laboratórios de informática, no ano de 2015 exclusivamente por professores e já no ano de 2016 por professores e bolsistas do projeto. Como forma complementar as aulas, foram utilizadas as plataformas URI e SPOJ Brasil [URI, 2016; SPOJ. 2016], as quais são repositórios de problemas de olimpíadas regionais, seletivas e demais competições da área. Durante o treinamento e com a inserção de listas de atividades dispostas nas plataformas, os alunos vivenciaram uma competição prévia com os demais participantes de todo o país.

Figura 2. Aplicação da OBI 2015 no IF Goiano – Campus Ceres

O material de divulgação do projeto foi elaborado pela equipe de treinamento, que também tem desenvolvido material didático próprio com base em informações coletadas durante os treinamentos, de forma especí ca ao ensino de programação para a OBI.

Ao nal de cada realização da OBI um questionário avaliativo foi enviado aos competidores para que pudessem opinar quanto a sua participação, e disposto um campo para que pudessem discorrer de forma subjetiva, conforme o depoimento de um dos estudantes da 1ª Série do Ensino Médio:

> "...Os alunos que zeram os treinamentos perceberam que com os treinamentos seus conhecimentos aumentaram acompanhado de suas notas, e isto os motivaram muito mais para persistirem na área das Ciências da Computação do que antes, pois esta ação de criar os treinamentos surtiu muitos efeitos, chamando assim suas atenções para seguirem suas carreiras rentes a esta área..."

Um dos alunos da 2ª Série do Ensino Médio comenta que sempre teve o intuito de participar do projeto, mesmo antes de ingressar no Ensino Médio:

> "Participar da Olimpíada Brasileira de Informática (OBI) sem dúvida foi uma das realizações mais grati cantes em minha vida. Desde muito novo me interessei pela área da programação. Acompanho as edições da OBI desde 2013, todavia antes de me ingressar no Instituto Federal Goiano Campus Ceres, não obtive a oportunidade de participar da mesma..."

Uma das estudantes da 3ª Série do Ensino Médio, participante da OBI em 2015 e 2016, comenta que de niu sua opção por curso na área de Tecnologia da Informação pela satisfação:

> "...Minhas experiências com a OBI foram essenciais para me ajudar na decisão do curso superior que desejo, ela também me auxiliou nas aulas de linguagem e tópicos, pois aprendi mais a lógica e as estruturas. Além, de me ajudar nas matérias técnicas ela também me auxiliou nas matérias do ensino médio como matemática, física e português, pois nos treinamentos sempre estava resolvendo exercícios que desenvolvia uma habilidade de interpretação dos problemas e nas resoluções precisava fazer cálculos matemáticos ou fórmulas de física."

5. Considerações nais

A participação de estudantes em olimpíadas cientí cas visa nivelar os conhecimentos em uma determinada área e aumenta o interesse dos mesmos em se aperfeiçoarem nos estudos.

A realização da OBI no Campus Ceres do IF Goiano além de ter obtido bons resultados nos últimos dois anos, como sendo a escola com o maior número de estudantes classi cados para a Fase 2 da Modalidade de Programação do Centro-Oeste, tem aumentado o interesse dos estudantes pelo curso e demonstrou que possuem capacidade de competir a nível nacional.

Devido as condições que são dispostas para a realização destas olimpíadas e pelo setor de tecnologia da informação ter necessidade de contratar muitos pro ssionais, a participação nestas olimpíadas tem aprimorado o currículo dos participantes para que assim possam se quali car para o mercado de trabalho e ao se formar conseguirem um emprego com maior facilidade.

Analisando os resultados de ensino-aprendizagem é possível veri car que a capacidade em adquirir conhecimento com algoritmos de programação é mais rápida e fácil com estudantes que possuam menor faixa etária e nenhum "vício" de programação.

Apesar da aplicação da OBI no Campus Ceres ter sido bem aceita pela comunidade acadêmica e ter trago bons resultados, faz-se necessário para os próximos anos um contato maior com a sociedade, abrindo assim turmas de treinamento para

estudantes de todas escolas de Ensino Médio de Ceres e região.

6. Referências Bibliográ cas

Boulic, R. and Renault, O. (1991) "3D Hierarchies for Animation", In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley & Sons ltd., England.

Brasscom, Brasil precisa de 750 mil novos pro ssionais de TI até 2020. Disponível em: <http://www.brasscom.org.br/brasscom/Portugues/detNoticia.php?codArea=6&codCategoria=8&codNoticia=1084> Acesso em: 02/08/2016.

Catho, Cresce o número de vagas de emprego na área de TI. 2016. Disponível em: <http://g1.globo.com/jornal-hoje/noticia/2015/07/cresce-o-numero-de-vagas-de-emprego-na-area-de-ti.html> Acesso em: 02/08/2016.

Instituições da Rede, Instituto Federal. 2016. Disponível em: <http://redefederal.mec.gov.br/?option=com_content&view=article&id=1001:unidades-da-rede> Acesso em: 04/10/2016

OBI, XVIII Olimpíada Brasileira de Informática. 2016. Disponível em: <http://olimpiada.ic.unicamp.br/info/geral> Acesso em: 04/10/2016

Pacheco, E; Rezende, C. (2009) "Institutos Federais Lei 11.892, de 29/12/2008: Comentários e Re exões", Editora IFRN, Brasília-DF.

Pacheco, E. (2010) "Os Institutos Federais: Uma revolução na educação pro ssional e tecnológica", Editora IFRN, Brasília-DF.

Rede Federal, Histórico. 2016. Disponível em: <http://redefederal.mec.gov.br/historico> Acesso em: 04/10/2016

Silva, D. L. M.; Costa, L. F. S.; Silva, M. A. A; Dantas, Ayla. Atraindo Alunos do Ensino Médio para a Computação: Uma Experiência Prática de Introdução a Programação utilizando Jogos e Python. In: XXII Simpósio Brasileiro de Informática na Educação, 2011, Aracaju. Anais do XXII Simpósio Brasileiro de Informática na Educação, 2011.

SPOJ, Sphere Online Judge. 2016. Disponível em: <http://br.spoj.com/> Acesso em: 04/10/2016

Sudol L. A. Deepening Students' Understanding of Algorithms: E ects of Problem Context and Feedback Regarding Algorithmic Abstraction. 2011. 27 f. School of Computer Sciente. Carnegie Mellon University. 2011.

URI, URI Online Judge. 2016. Disponível em: <https://www.urionlinejudge.com.br/judge/pt/> Acesso em: 04/10/2016

Index of Authors