Anais da I Escola Regional de Sistemas de Informação do Rio de Janeiro

Proceedings of the 1st Regional School on Information Systems of Rio de Janeiro

Niterói, 4 a 7 de novembro de 2014

Organidores

Claudia Capelli - Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

José Viterbo - Universidade Federal Fluminense (UFF)

Promoção

Sociedade Brasileira de Computação

Editores

Claudia Capelli (Universidade Federal do Estado do Rio de Janeiro)
José Viterbo (Universidade Federal Fluminense)

Título – Anais da I Escola Regional de Sistemas de Informação do Rio de Janeiro
Local – Niterói/RJ, de 4 a 7 de Novembro de 2014
Ano de Publicação – 2014
Edição – 1ª
Editora – Sociedade Brasileira de Computação - SBC
Organizadores – Cláudia Cappelli (UNIRIO)
José Viterbo (UFF)

ISBN: 978-85-7669-294-2 (online)

Apresentação

A sociedade vive um momento em que a tecnologia cada vez mais aumenta as possibilidades de se partilhar as funções cognitivas dos indivíduos através do suporte eletrônico. Organizações, Instituições e Empresas (OIEs), são diretamente afetadas por esta nova realidade, e requerem, necessariamente, a formação de profissionais que tenham condições de assumir um papel de agente transformador do mercado, sendo capazes de provocar mudanças através da incorporação de novas tecnologias da informação na solução dos problemas.

É urgente a formação de profissionais com visão interdisciplinar, global, crítica, empreendedora e humanística que possam viabilizar a busca por soluções computacionais para problemas complexos de áreas diversas, considerando não somente questões técnicas relativas ao processamento da informação, mas também todo o contexto humanístico que abriga o problema em questão, é com essa perspectiva que a se insere a proposta desse evento.

I Escola Regional de Sistemas de Informação do Rio de Janeiro (ERSI-RJ 2014), é um evento que teve como objetivo principal reunir estudantes, professores, pesquisadores e profissionais de Sistemas de Informação para promover discussões sobre temas relacionados a esta área. A programação do primeiro ERSI-RJ contou com diversas atividades que se destacam pelo seu alto nível técnico-científico e que cobrem diferentes aspectos, incluindo palestras convidadas com especialistas nacionais, painéis com representantes da academia, indústria e governo, sessões técnicas para a apresentação dos artigos científicos selecionados e minicursos e tutoriais em diversas áreas.

Leonardo Cruz

Coordenador Geral

Comitê Organizador do Evento

Coordenação Geral
Leonardo Cruz (UFF)
Rodrigo Salvador (UFF)

Coordenação de Programa
Claudia Cappelli (UNIRIO)
José Viterbo (UFF)

Coordenação de Minicursos
Carlos Eduardo Melo (UFRRJ)
Ilaim Costa (UFF)

Coordenação de Tutoriais
Diego Brandão (CEFET-RJ)
Henrique Viana (CEFET-RJ)

Comitê de Seleção de Tutoriais
Alexandre Sena (UERJ e UNILASALLE)
Bruno Nascimento (UFRJ e FEUC-RJ)
Diego N. Brandão (CEFET-RJ)
Francisco Eduardo Cirto (CEFET-RJ)
Henrique Viana CEFET-RJ
João Vinicius Thompson (UFF)
Vinícius Forte (COPPETEC-UFRJ e UniMSB)
Vinicius Petrucci (Universidade de Michigan-Ann Arbor/USA)
Rafael Augusto de Melo (UFBA)
Ueverton dos Santos Souza (CEFET-RJ)

Comitê de Seleção de Melhor Artigo Técnico
Anselmo Montenegro (UFF)
Daniela Trevisan (UFF)
Marcos Kalinowsky (UFF)

Comitê de Programa

Alessandro Copetti (UFF)
Alexandre Correa (UNIRIO)
Alexandre Sena (UERJ e UNILASALLE)
Alexandre Veloso (UDESC)
Aline Paes (UFF)
Asterio Tanaka (UNIRIO)
Carolina Felicíssimo (PUC-Rio)
Carlos Eduardo Melo (UFRRJ)
Claudia Cappelli (UNIRIO)
Cristiano Maciel (UFMT)
Daniel Paiva (UTFPR)
Daniela Trevisan (UFF)
Diego Brandão (CEFET-RJ)
Ecivaldo Matos (UFBA)
Fabiana Mendes (UnB)
Fernanda Baião (UNIRIO)
Flavia Bernardini (UFF)
Flávia Maria Santoro (UNIRIO)
Geiza Hamazaki (UNIRIO)
Gleison Santos (UNIRIO)
Ilaim Costa Junior (UFF)
Isabel Cafezeiro (UFF)
Jonice Oliveira (UFRJ)
José Ricardo Cereja (UNIRIO)
Jose Viterbo (UFF)
Leonardo Cruz (UFF)
Luciana Salgado (UFF)
Reinaldo Alvares (UNISUAM)
Renata Mendes Araujo (UNIRIO)
Rodrigo Salvador Monteiro (UFF)
Rosangela Lima (UFF)
Sergio Crespo (UFF)
Sean W. M. Siqueira (UNIRIO)
Silvia Amelia Bim (UTFPR)
Simone Bacellar Leal Ferreira (UNIRIO)
Victor Almeida (Petrobras)

ARTIGOS TÉCNICOS

Sessão Técnica 1 – Ensino de SI

Sessão Técnica 2 – Aplicações de SI (1)

Sessão Técnica 3 – Melhores artigos

Sessão Técnica 4 – Arquiteturas de SI

Sessão Técnica 5 – Aplicações de SI (2)

PET-SI: Em Busca de Formação Diferenciada Para Discentes de SI através da Fábrica de Software Baseada em Métodos Ágeis

Felipe Calé[1,2], Jeferson Leonardo[1,2], Gustavo Sucupira[1,2]

Renan Toyoyama[1,2], Sérgio Manuel Serra da Serra[1,2,3]

[1]Universidade Federal Rural do Rio de Janeiro/UFRRJ
[2] Programa PET Sistemas de Informação - PET-SI/UFRRJ
[3] Programa Pós Graduação em Modelagem Matemática e Computacional/UFRRJ

{ felipe, jeferson, gustavo, toyoyama, serra }@pet-si.ufrrj.br

Resumo. *O PET-SI é um programa de Educação Tutorial que tem como objetivo capacitar alunos em situação de fragilidade financeira para que trabalhem em equipe, despertando o interesse a adoção de boas práticas em TI e desenvolvimento de ações voltadas para a inovação tecnológica em computação. Este trabalho apresenta a metodologia FSMA desenvolvida pelo grupo PET-SI e discute sua importância para o curso de Sistemas de Informação da UFRRJ. O trabalho também apresenta os principais produtos e benefícios gerados pelo programa e a relação do ganho de desempenho acadêmico dos alunos ao longo dos primeiros meses de funcionamento.*

1. INTRODUÇÃO

A história e a geografia da Baixada Fluminense estão ligadas à UFRRJ, ela possui fronteira com municípios com milhares de habitantes, maioria de baixa renda. Os municípios têm baixo IDH e não possuem empresas de tecnologia da informação (TI) nem tradição no setor de TI. O curso de Sistemas de Informação (SI) iniciou em 2010-1, hoje possui 16 professores, 112 alunos ativos, 78,3% da rede pública, e 27 inativos. O Programa de Ensino Tutorial de Sistemas de Informação (PET-SI) intitula-se "A Tecnologia da Informação como agente de transformação social" e alinha-se com PDI da universidade. Tem como objetivos ampliar a oferta de profissionais de TI e preparar egressos para serem eles próprios agentes de transformação social nas suas comunidades. O PET-SI (http://r1.ufrrj.br/petsi/) tem como função social mitigar efeitos da reprovação/evasão/retenção e melhorar a formação do profissional ampliando a capacidade de inovar, planejar e gerenciar serviços e recursos da TI, tornar os alunos aptos a empreender e liderar equipes. Hoje o PET-SI é composto por 12 alunos bolsistas (todos em situação de fragilidade financeira) e 1 tutor, tendo como público-alvo 120 alunos do curso de SI e nível médio das escolas dos municípios do entorno. Atualmente o PET-SI mantém parcerias com a *Google* (através do programa Google Apps for Education), além da empresa Junior *Ceres Jr.* na capacitação das ferramentas do Google em mini-cursos ministrados pelos petianos do PET-SI. Recentemente estabeleceu parcerias com a *Amazon* (parceria educacional sobre treinamento em *Cloud Computing*), *CRBD-UFRJ* (parcerias para disseminar o conhecimento em *Big Data)*.

O PET-SI possui três pilares de ação: i) Ações de ensino, extensão e pesquisa, mediante a formação de grupos de aprendizagem tutorial voltados para criar softwares e cursos de capacitação desenvolvidos por alunos para as comunidades Universitária e da Baixada; ii) Elevar a qualidade acadêmica dos alunos através da formulação de ações de pesquisa e extensionistas voltadas para as comunidades humanas e industriais do entorno, estimulando projetos de inovação tecnológica que adotem políticas de sustentabilidade e geração de renda através de práticas interdisciplinares em computação e reciclagem; iii) Estimular o senso crítico, promover ação pautada na ética, cidadania e na função social proporcionada pela educação superior. O PET-SI alinha a teoria à prática e aprofunda os conhecimentos adquiridos em sala, desenvolveu a metodologia *Fábrica De Software Baseada Em Métodos Ágeis* (FSMA) (Cruz et al, 2013 e 2014) de apoio à aprendizagem a resolução de situação-problema de cunho socioambiental e baseia em SW-livres, nuvens de computadores e de técnicas de gestão de Big Data.

O PET-SI desenvolveu a nova metodologia que se difere do conceito de fábrica de software tradicional (Pressman, 2011). A FSMA se alinha com a escola de pensamento proposta por Vygostky (2007) onde a educação é considerada como fonte de desenvolvimento. Aqui se defende que o desenvolvimento ágil de software deva fazer parte do processo de ensino-aprendizado (PEA) focada no aluno, buscando-se a melhoria continuada de forma continuada e iterativa. Portanto, é preciso treinar e manter times de alunos trabalhando juntos, em um mesmo domínio e próximos ao cliente para gerar eficiência e aprendizado na prática, sem abrir mão da qualidade, controlando os custos de desenvolvimento, tempo de entrega do produto.

2. METODOLOGIA

O PET-SI tem como método capacitar os alunos para trabalhem em equipe, despertar o interesse a adoção de boas práticas em TI e desenvolvimento de ações voltadas para a inovação tecnológica. A FSMA aborda conceitos de Métodos Ágeis, *Exteme Programming* e PEA que tem sido caracterizado por diferentes formas que vão desde a ênfase no papel do professor como difusor de conhecimento, até as concepções atuais que concebem o processo de ensino-aprendizagem com um todo integrado que destaca o papel do aluno (Coll & Mira, 1996, Perrenoud, 2000, Vygostky, 2007). Os experimentos em desenvolvimento de software e treinamentos são realizados de modo colaborativo, criando um espaço de referência, onde as funcionalidades dos sistemas são estruturadas em função dos saberes previamente adquiridos ao longo do curso de graduação pelos petianos e das necessidades das comunidades de alunos do curso de graduação.

Os experimentos foram executados entre Março de 2013 e Setembro de 2014, contando com a participação de 15 petianos (42% mulheres e 58% homens). Todos os alunos envolvidos cursavam o último período do ciclo básico do curso de Sistema de Informação da UFRRJ. Semanalmente, ocorriam encontros com duração de aproximadamente duas horas com todos componentes. A cada encontro o tutor avaliava (individual e coletivamente): (i) o desempenho dos alunos e suas tarefas de desenvolvimento; (ii) o cronograma dos projetos; (iii) as funcionalidades implementadas em cada sistema. No decorrer da semana, os alunos se comunicavam através sistemas online (por exemplo, Hangout do Google ou Skype) reportando seus progressos.

3. RESULTADOS E DISCUSSÃO

Os experimentos desenvolvidos pelos alunos foram compostos pelas seguintes etapas (i) análise e levantamento de requisitos dos (dois) sistemas; (ii) definição da arquitetura dos sistemas; (iii) implementação; (iv) testes de funcionalidades e (v) implantação das soluções na infraestrutura de servidores Web já disponíveis na Universidade. O PET-SI desde a sua formação tem estimulado a participação em eventos científicos locais e nacionais além da publicação dos resultados dos experimentos sob a forma de trabalhos acadêmicos que dão visibilidade às ações e que também possibilitam aos petianos uma experiência acadêmica diferenciada. A Tabela 1 resume os trabalhos apresentados desde a fundação do grupo PET-SI.

Tabela 1. Trabalhos publicados pelo PET-SI desde Setembro de 2013

Título do trabalho	Evento/Local/Ano
Primeiros estudos empíricos sobre uma fábrica de software baseada em métodos ágeis no PET-SI/UFRRJ	ENAPET (Encontro Nacional dos Grupos Pet) - UFPE, Recife-PE - 2013
Primeiros estudos em computação em nuvens no PET-SI apoiados pelo *google apps for education*	I RAIC da UFRRJ - UFRRJ, Seropédica-RJ - 2013
Primeiros estudos empíricos sobre uma fábrica de software baseada em métodos ágeis no PET-SI/UFRRJ	I RAIC da UFRRJ - UFRRJ, Seropédica-RJ - 2013
Avaliação do comportamento eletrônico dos usuários do Sítio do SudestePet2014 através do *Google Analytics* e Facebook	Sudeste PET - UFRRJ, Seropédica-RJ - 2014
Aperfeiçoamento da Fábrica de Software do PET-SI: Estudo de Caso no SudestePET 2014	Sudeste PET - UFRRJ, Seropédica-RJ - 2014
Combinando dados de *clickstream* e análise de redes sociais para identificação do comportamento eletrônico dos petianos da região sudeste	ENAPET - UFSM, Santa Maria-RS - 2014
Minicurso de *Google Apps*: Relato de uma experiência extensionista do PET-SI na UFRRJ	II RAIC da UFRRJ - UFRRJ, Seropédica-RJ - 2014
PET-SI na internet: Análise de navegação baseada em *web anaytics* e redes sociais	II RAIC da UFRRJ - UFRRJ, Seropédica-RJ - 20124

Ao longo dos dois últimos anos, observamos que os petianos melhoram o seu desempenho acadêmico após ingressarem no PET-SI, isso se deve à grande demanda por pesquisas extra-classe,

leituras e atividades desenvolvidas por participantes, além do acompanhamento realizado pelo tutor que provê atenção particularizada. As responsabilidades são distribuídas individualmente pelo tutor, fortalecendo a maturidade individual, formação cidadã e convivência do grupo. A Figura 1 exibe a evolução média do coeficiente de rendimento (CR) dos petianos (antes e depois de ingressarem no PET-SI). O CR médio dos alunos não petianos é aproximadamente 5.0. Destaca-se que todos os alunos do PET-SI estavam em risco de evasão e a eles foram concedidas de bolsas de Iniciação Científica. Dentre os membros do grupo, 13,5% tiveram a oportunidade de serem contemplados com bolsas extras da CAPES/CNPQ para programa CSF.

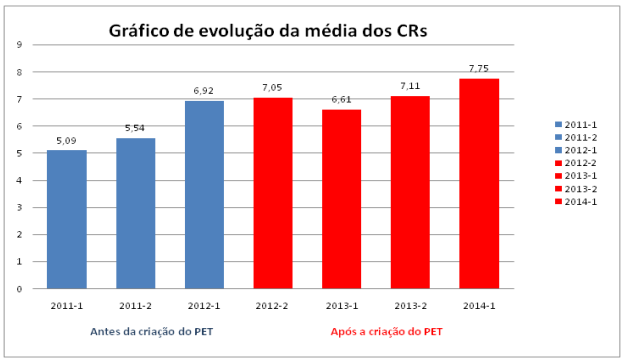


Figura 1. Gráfico com a evolução do CR médio dos alunos do PET-SI (N=15 alunos).

4. CONCLUSÃO

O PET-SI propiciou até o momento aos alunos do curso de SI a oportunidade de expandir e consorciar o conhecimento teórico com atividades práticas de caráter científico e extensionista. Até o momento houve aumento no CR médio, redução do número de reprovações e ausência de evasão entre petianos. A UFRRJ e as comunidades do entorno se beneficiam dos projetos de ensino-pesquisa-extensão.

AGRADECIMENTOS

Os autores agradecem à FAPERJ e ao FNDE/MEC pelas bolsas concedidas e pelo apoio financeiro às pesquisas.

REFERÊNCIAS

Cruz, S.M.S. et al (2013) "Relato De Um Experimento Piloto De Uma Fábrica De Software Baseada Em Métodos Ágeis". Anais do XVII ENAPET, Recife-PE.

Cruz, S.M.S, Sucupira, G. Leonardo, J. (2014) "Aperfeiçoamento Da Fábrica De Software Do PET-SI: Estudo De Caso No Sudeste Pet 2014″. Anais do XIV SUDESTEPET, Seropédica-RJ.

Coll, C. e Miras, M. (1996) "A Representação Mútua Professor/Aluno e Suas Repercussões Sobre o Ensino e a Aprendizagem. Desenvolvimento psicológico e educação: psicologia da educação", ArtMed. p.265-280.

Perrenoud, P. (2000) "Dez Novas Competências Para Ensinar". Artes Médicas Sul. 192p.

Pressman, R., (2011) "Engenharia de Software – Uma Abordagem Profissional" McGraw-Hill. 780p.

Vygostky, L. S. (2007) "Formação social da mente". Martins Fontes, 2ª Edição.

Projeto Interagir de Capacitação em Informática: Construção de saberes sobre o uso das Tecnologias de Informação e Comunicação no âmbito das graduações da UFF

Jean P. Campos[1], Isabel Cafezeiro[1], Rosangela L. Lima[1]

[1]Instituto de Computação – Universidade Federal Fluminense (UFF) - Niterói, RJ

isabel@dcc.ic.uff.br, lima@dcc.ic.uff.br, jeancampos@iduff.br

Abstract. *This article discusses a project that seeks the interaction between the educational and computing areas, proposing the implementation of a knowledge construction network mediated by technology. Through the cooperation of under-graduated students, staff and professors, the Projeto Interagir aims to turn possible challenges and limitations present on the access to digital media and technological tools into factors of mutual interest, aiming the construction of knowledge that may contribute to the academic and personal development.*

Resumo. *Este artigo discute um projeto que busca a interação entre as áreas da educação e computação, propondo a efetivação de uma rede de construção de saberes mediada pela tecnologia. Através da cooperação de graduandos, funcionários e professores, o Projeto Interagir visa transformar eventuais dificuldades e limitações existentes no acesso aos meios digitais e ferramentas tecnológicas em fatores de interesse mútuo, com vistas na construção de conhecimentos que possam contribuir na formação acadêmica e pessoal.*

1. O contexto da aplicação de projeto e a proposta

A tecnologia como mediadora de processos de ensino se apresenta como possibilidade de inovação no processo de ensino e de aprendizagem das mais diferentes áreas do conhecimento.

> "[...] para que o processo ensino-aprendizagem obtenha êxito é importante criar condições a fim de que professores e alunos estejam próximos, mesmo que ainda distantes. Ao professor cabe motivar o aluno e acompanhar o processo; ao aluno cabe ser mais autônomo e proativo; e à tecnologia cabe oferecer soluções e ambientes mediadores deste processo." (LOPES, 2007, p. 103).

Computadores ligados a internet com o apoio de *datashow* são nos dias de hoje o quadro negro e o giz de tempos atrás, mas a sua utilização com o aproveitamento de todas as potencialidades que surgem com a presença da Internet não parece ser o *modus operandi* adotado em grande parte das salas de aula do ensino superior. O que se vê com frequência é um grande contingente de alunos que não têm acesso nem familiaridade com essa tecnologia. Enquanto o professor reproduz antigas práticas mesmo com o uso do computador e do *datashow,* o estudante, que no cotidiano convive com a tecnologia das redes sociais, em sala de aula fica impedido de aprender a buscar informações, elaborar e gerenciar conteúdos, que são práticas fundamentais para a construção do conhecimento significativo tão importante para a sua formação acadêmica.

> "Conceitos como "alfabetização digital", "alfabetização informacional" ou "alfabetização em mídia" acenam para cenários que requerem habilidade crítica e autocrítica no manejo da informação e do conhecimento." (Demo, 2011).

Diante deste cenário a Universidade Federal Fluminense (UFF), por meio de uma parceria entre a Pró-Reitoria de Assuntos Estudantis (PROAES) e o Instituto de Computação (IC), vislumbrou a oportunidade de intervir de forma efetiva na problemática apontada a partir do oferecimento de suporte a estudantes de graduação no uso das tecnologias de informação e comunicação (TIC). Foi criado assim o Programa Interagir de Capacitação em Informática na UFF visando suprir as dificuldades mais comuns relacionadas ao uso das TIC apresentadas por alunos graduandos no cumprimento de suas atividades acadêmicas. Este programa que teve início em 2013, atualmente se encontra em sua quarta edição.

Este projeto se distingue dos usuais programas de Inclusão Digital cuja proposta costuma ser a capacitação técnica nas ferramentas básicas. O Projeto Interagir difere deste modelo em pelo menos quatro sentidos: (1) Quanto ao conteúdo que oferece: o programa não parte de um conteúdo preestabelecido, ao contrário, ele se molda à dinâmica das diversas graduações. (2) Quanto à participação do aluno atendido: o aluno tem participação ativa na configuração do curso, estabelecendo em entendimento com os tutores e professores orientadores desde a escolha dos softwares até abrangência dos assuntos abordados. (3) Quanto à participação do tutor: a equipe de tutores concebe, estrutura e ministra seus cursos a partir das solicitações apresentadas pelas turmas, adotando para isto um ambiente virtual de aprendizagem, o Portal Interagir (www.interagir.uff.br). Os tutores responsabilizam-se por todas as etapas, da divulgação ao encerramento, com o acompanhamento dos professores orientadores. (4) Quanto à avaliação continuada: o ambiente virtual de aprendizagem permite a interação permanente entre todos que participam do curso, viabilizando a avaliação continuada do aprendizado, o que fornece diretrizes para novos planejamentos.

2. Objetivo do Projeto Interagir

As ferramentas tecnológicas são viabilizadoras da cooperação no processo de construção de saberes. A partir dessa consideração optou-se por adotar uma metodologia que buscou reunir estes alunos de modo que a troca de saberes mediada pela tecnologia pudesse contribuir fortemente no aprendizado de todos os envolvidos na proposta pedagógica. A inovação pretendida partiu de uma concepção cujo foco teria origem na necessidade dos alunos e não em um conteúdo pré-determinado. Na Tabela 1 apresenta-se um exemplo da diversidade de alunos em termos das áreas de conhecimento no período de 2014.1 e 2014.2, nas duas últimas edições.

Tabela 1. Diversidade de cursos no Projeto Interagir

Áreas do Conhecimento	Número de alunos
Ciências Exatas e Engenharias	188
Ciências Sociais e Humanas	172
Ciências Biológicas e da Saúde	70

Dada a solicitação da PROAES a partir da constatação da dificuldade dos graduandos em acompanhar as atividades de seus cursos, o Instituto de Computação se empenhou em propor um mecanismo que ao mesmo tempo atendesse a estes graduandos e também contribuísse na formação dos alunos do próprio instituto. A iniciativa visou delegar a eles a

autonomia e responsabilidade na condução do projeto, bem como discussão e esclarecimento dos fundamentos pedagógicos. Assim,

> "a estratégia pedagógica é baseada na participação ativa de cada aluno em um processo de *co-construção*, ou seja, na elaboração conjunta de conteúdos, tecnologias e avaliações a partir do diálogo com os alunos em termos de suas próprias demandas (ou das demandas de seus cursos), o que possibilita ao mesmo tempo identificar e sobrepor dificuldades e atender dificuldades localizadas." (CAMPOS, CAFEZEIRO, LIMA, 2014, p. 110).

A adoção dessa estratégia, à primeira vista, pode causar estranheza, mas na prática envolve alunos, professores e tutores de modo surpreendente possibilitando a eficácia e a aceitação do projeto, como se pode constatar no depoimento da estudante de Farmácia.

> "Eu trabalho muito com o editor de textos, com o *excel* e com o *power point*. Nas 18 horas de curso que tivemos aprendemos mais sobre o 'pacote office' (libreoffice - *software* livre), que inclusive eu atualmente utilizo, mas eu só conheci esse software no Projeto Interagir. Aprendemos sobre a Computação em Nuvem, que eu ainda não uso, por exemplo Google drive, mas notei que é bastante útil e vantajoso." (PROAES, 2014).

Na perspectiva da cooperação, o conceito de software livre representou um pilar fundamental na formação de uma cultura baseada no compartilhamento e adaptação, uma vez que a demanda dos alunos em sala é fator decisivo no planejamento do escopo das aulas. A expertise dos tutores fornece um leque de possibilidades relacionadas às funcionalidades e facilidades características do software livre, o que propicia a criação de uma cultura de tecnologia que vai além das ofertas proprietárias. Estas últimas, uma vez que não tornam acessíveis seus mecanismos de construção posicionam-se no contra sentido da colaboração, terminado por constituir-se um fator fortemente responsável pela dependência tecnológica.

No que se refere à formação profissional, a proposta vai além dos estudantes que se inscrevem no curso. Os graduandos como tutores dos cursos, se tornam responsáveis por todas as etapas da construção de cada edição, desde a concepção dos tópicos, negociação com a turma sobre o fluxo da aula, viabilização dos recursos, seleção das ferramentas, até a autoavaliação e avaliação da turma. Esta prática adquirida lhes dá experiência na atuação como docentes, pois participam de todas as etapas de planejamento, monitoramento, gestão e execução do curso. Por meio da participação destes alunos tutores pretende-se fomentar seus saberes através da concepção e execução de tópicos em um processo de co-criação, além de agregar a eles conhecimentos sobre prática didática.

3. Ampliação do projeto e infraestrutura de mediação tecnológica

A partir da experiência adquirida na primeira edição, o Projeto foi ampliado para outras unidades da UFF, desta forma mais uma atividade coube aos tutores. Esta atividade consistiu na sua atuação como multiplicadores qualificando outros alunos que se tornaram tutores em suas unidades propiciando a disseminação do projeto. Visitas às unidades fora de Niterói foram incorporadas às atividades dos tutores para garantir a boa funcionalidade do projeto nas unidades do interior. Assim, a partir do sucesso das primeiras edições, o projeto passou a abranger também as unidades da UFF localizadas no interior do estado do Rio de Janeiro, onde as dificuldades com o uso das tecnologias são mais evidentes.

Na edição atual o projeto inicial foi adaptado a um modelo de tópicos e ampliado estendendo-se aos pólos de Rio das Ostras e Volta Redonda, nos quais se replica a mesma

proposta em total integração com a unidade Niterói. A integração e comunicação entre estas ações se faz por meio de encontros presenciais e pela mediação de um ambiente virtual de aprendizagem do Portal Interagir (www.interagir.uff.br). Baseado no MOODLE, o portal serve não somente como um fator de aproximação entre localidades fisicamente distantes, mas também como dinamizador do diálogo, ferramenta de treinamento na cultura do projeto, instrumento onde se pode ter acesso a conteúdos de tópicos anteriormente implementados, além de permitir o compartilhamento de soluções experimentadas com as demais unidades.

4. Conclusão

Delors (1999) explica que a educação deve se organizar em torno de quatro aprendizagens, às quais ele dá o nome de "pilares do conhecimento". A estratégia educativa deste projeto está em sintonia direta com estes quatro pilares: "*aprender a conhecer*, isto é, adquirir os instrumentos da compreensão", o que neste projeto se manifesta pela busca curiosa dos recursos tecnológicos. "*Aprender a fazer*, para agir sobre o meio envolvente", aqui presente na construção criativa através dos recursos tecnológicos. "A*prender a viver juntos*, a fim de participar e cooperar com os outros em todas as atividades humanas", uma vez que há a demanda por colaboração, e finalmente, "A*prender a ser*, via essencial que integra a três precedentes", pilar abraçado pela autonomia estimulada no projeto. Na utilização da tecnologia como instrumento transformador da sua própria realidade, em função de suas próprias demandas e expressões criativas foi possível tornar o aprendizado efetivo. O que ficou evidente no desenvolvimento do projeto foi a certeza de que no contexto educativo atual não se pode desconsiderar a tecnologia como ferramental imprescindível para a inclusão de saberes e competências relacionadas às TIC.

5. Referências

Campos, J., Cafezeiro, I., Lima, R., (2014). Programa Interagir de capacitação em informática. In *LibreOffice Magazine*, p. 109 - 113. Ed. 11, Junho.

Delors, J. (1999) "Os quatro pilares da educação", In: *Educação: um tesouro a descobrir*. São Paulo: Cortezo. p. 89-102.

Demo, P. (2011) "Olhar do educador e novas tecnologias", B. Téc. Senac: a R. Educ. Prof., Rio de Janeiro, v.37, n.2, p.15-26. http://www.senac.br/BTS/372/artigo2.pdf, Outubro.

Lopes, M. S. S. (2007) "O professor diante das Tecnologias de Informação e Comunicação em EAD", In: Formação Docente: diferentes percursos. Rio de Janeiro.

PROAES. (2014) "Projeto Interagir conquista estudantes e funcionários", http://www.proaes.uff.br/projeto-interagir-conquista-estudantes-e-funcionarios, Setembro.

Combinando Metodologias Ágeis para Execução de Projetos de Software Acadêmicos

Narallynne M. Araújo, Flavius L. Gorgônio, Karliane M. O. Vale

Universidade Federal do Rio Grande do Norte (UFRN)
Laboratório de Inteligência Computacional Aplicada a Negócios (LABICAN)
Rua Joaquim Gregório, S/N, Penedo, CEP 59300-000 – Caicó – RN – Brasil

{narallynne, flgorgonio, karlianev}@gmail.com

Abstract. *Considering the limitations of applicability to commercial software development methodologies in academic scope, this paper proposes guidelines to the development of a new methodology based on XP, RUP and Scrum methodologies that aim to add on teaching and academic practices, an approach based on the outside university environment.*

Resumo. *Considerando as limitações de aplicabilidade de metodologias de desenvolvimento de software comerciais em âmbito acadêmico, este trabalho propõe diretrizes para o desenvolvimento de uma nova metodologia baseada nas metodologias XP, RUP e Scrum que busca agregar ao ensino e à prática acadêmica, uma abordagem baseada no que acontece fora do ambiente universitário.*

1. Introdução

Desde o surgimento da filosofia do Manifesto Ágil em 2001, as metodologias ágeis de desenvolvimento de software vêm sendo comumente utilizadas no mercado de trabalho de tecnologia em grandes e pequenas empresas desenvolvedoras [Ludvig e Reinert 2007]. Uma vez que o aprendizado da Engenharia de Software (ES) e das metodologias de desenvolvimento ocorre na Universidade, é perceptível a distinção entre os projetos de software desenvolvidos em âmbito acadêmico e no mercado de trabalho [Garcia *et al.* 2004] e tais diferenças quanto à formação das equipes, experiência de trabalho, tempo de dedicação ao projeto e o seu escopo são facilmente identificadas pelos egressos ao chegarem no mercado.

Com o intuito de amenizar essa desigualdade, as instituições de ensino superior buscam cada vez mais adaptar o ensino dessas metodologias ao ambiente acadêmico, desenvolvendo projetos de disciplinas com mais qualidade [Rodrigues e Estrela 2012]. Por exemplo, no curso de Bacharelado em Sistemas de Informação (BSI) da Universidade Federal do Rio Grande do Norte (UFRN), as metodologias ágeis são utilizadas em projetos das disciplinas da área de ES com adaptações, pois a sua aplicabilidade voltada para o mercado de trabalho não se adequa à realidade acadêmica e vice-versa.

Diante das divergências em encontrar uma metodologia que seja didática e, ao mesmo tempo, atenda às necessidades comuns aos dois âmbitos, vários autores propõem adaptações a partir da análise de metodologias comumente utilizadas no mercado, tais como XP, RUP e Scrum [Garcia et al.2004], [Rodrigues e Estrela 2012]. Neste trabalho,

pretende-se propor uma solução mais simples, explorando deficiências já identificadas em trabalhos anteriores, porém capaz de agregar ao ambiente acadêmico um conjunto de práticas voltadas ao mercado, de forma que esse procedimento seja, ao mesmo tempo, didático e aplicado à realidade dos projetos acadêmicos.

2. Aplicação de metodologias ágeis no mercado e em ambiente acadêmico

O *Rational Unified Process* (RUP), ou simplesmente Processo Unificado, representa a unificação de um conjunto de metodologias tradicionalmente utilizadas até meados dos anos 90. A metodologia possui um processo baseado em componentes (iniciação, elaboração, implementação e implantação), onde o produto de software deve ser desenvolvido mediante a unificação desses componentes, daí o seu nome. A gestão do seu ciclo de vida é capaz de estabelecer a visão do projeto, definir e validar a arquitetura planejada, gerenciar recursos e testar produtos entregues ao cliente. Com isso, busca-se garantir que a produção de software tenha alta qualidade e que esteja de acordo com a necessidade dos usuários finais [Tsui, 2013], [Sbrocco e Macedo, 2012].

Para Rodrigues e Estrela (2012), o RUP possui um grande número de papéis, atividades e artefatos gerados, considerado muito extenso e complexo para suas exigências, o que dificulta a sua utilização em ambiente acadêmico. Além disso, não produz um produto capaz de atender às necessidades dos clientes, pois o RUP considera que o desenvolvedor precisa ter experiência nos conceitos e práticas de ES, o que nem sempre mostra-se verdade em equipes acadêmicas [Garcia *et al*.2004].

Beck (2004) propõe a metodologia *Extremme Programming* (XP), encorajando sua aplicabilidade em equipes pequenas que atuem em um mesmo local de trabalho, como forma de fomentar a comunicação [Tsui, 2013]. A metodologia sugere a criação de documentações que sejam rigorosamente necessárias, como as linhas de código e os testes de unidade usados como artefatos. Apresenta ainda alguns valores como forma de conduzir a equipe durante o processo de desenvolvimento, como a simplicidade das suas tarefas que vão desde a produção de código simples e questões referentes ao design, requisitos ou testes, onde nenhuma função desnecessária possa ser levada em consideração [Soares, 2004], [Sbrocco e Macedo, 2012].

Para Garcia *et al.* (2004), o uso do XP em ambiente acadêmico requer uma equipe de desenvolvimento experiente, o que deixa muitas práticas em aberto, uma vez que equipes formadas por alunos de graduação possuem pouca maturidade e conhecimento. Para Rodrigues e Estrela (2012), outra dificuldade na utilização dessa metodologia está no uso da programação em pares, rigorosamente seguido em sua utilização, uma vez que os alunos possuem dificuldade de trabalhar em duplas fora da sala de aula.

A metodologia Scrum [Schwaber e Sutherland 2011] possui características que abordam processos de maneira empírica, iterativa e incremental, referenciando o foco funcional e a capacidade de responder positivamente às mudanças que surgem durante o processo de desenvolvimento de software [Pham 2011]. Seu processo é orientado por ciclos de tempo pré-estabelecidos (*Sprints*), onde as iterações de trabalho são realizadas. A Scrum estabelece papéis (*Product Owner*, *Scrum Master* e *Team*) e cerimônias (*Daily Scrum*, *Sprint Review*, *Sprint Planning Meeting* e *Sprint Retrospective*), como artefatos

principais da sua teoria. Além de documentações unicamente necessárias (*Product Backlog, Sprint Backlog, Burndown Chart* e *Task Board*).

A Scrum possui um processo simples e enxuto, onde os membros da equipe do projeto interagem diariamente entre si. Porém, tais características são consideradas desafiadoras na tentativa da sua aplicação na academia por duas razões: a) devido aos perfis e responsabilidades distintas entre uma equipe formada por estudantes de graduação e outra equipe formada por pessoas no mercado de trabalho; b) no processo de aprendizado, uma vez que é necessário tentar unir estudo e experiência de mercado, necessitando de um processo didático e sob medida [Rodrigues e Estrela 2012]. Ludvig e Reinert (2007) afirmam ainda que a Scrum é voltada para a gestão do processo de desenvolvimento de software, enfatizando os aspectos de gerência e da organização da equipe, direcionada para pessoas que já tenham um grau de conhecimento em projetos de software e que queiram adaptá-lo às suas necessidades.

A metodologia acadêmica *easYProcess* (YP), derivada de metodologias utilizadas no mercado, tais como XP e RUP, possui seu foco unicamente acadêmico e com poucos trabalhos que a referenciam, sem nenhuma comprovação de que ela seja, de fato, aplicada no mercado. O fluxo do seu processo é centrado na produção de documentos que compreendem a definição dos papéis da equipe, conversa com o cliente, análise e planejamento. Por fim, são realizadas as implementações, seguidas da entrega e acompanhamento do projeto de software [Garcia *et al*.2004].

3. Proposta de diretrizes para o desenvolvimento de uma nova metodologia ágil para uso em ambiente acadêmico

Este trabalho tem como objetivo propor diretrizes para a integração de três metodologias ágeis: XP, RUP e Scrum, buscando agregar um aprendizado didático e funcional à realidade dos projetos acadêmicos. Com isso, almeja-se incorporar à prática de desenvolvimento de software na academia, um conjunto de técnicas voltadas ao que comumente acontece no mercado de tecnologia.

Neste trabalho, em desenvolvimento, os autores analisam o uso dessas metodologias tanto em trabalhos publicados na literatura quanto em projetos acadêmicos desenvolvidos no curso de BSI da UFRN, incluindo sua aplicabilidade em disciplinas das áreas de engenharia de software e gestão de projetos. Em uma autoavaliação realizada nessas disciplinas, constatou-se que a falta de experiência no uso de metodologias e a conciliação de horário da equipe são fatores que dificultam a prática acadêmica. Além disso, está sendo investigada a utilização dessas metodologias ágeis nas *software houses* da região do Seridó Potiguar, a fim de observar as adaptações do mercado de trabalho em seu uso cotidiano. Resultados preliminares indicam que cerca de 75% das empresas visitadas utilizam metodologias ágeis, principalmente baseadas no Scrum.

Por meio desse estudo e análise, os autores estão desenvolvendo um fluxo de processo adaptado e baseado nas três metodologias, onde os resultados iniciais já estão sendo experimentados em atividades do curso de BSI da UFRN. Com isso, pretende-se utilizar as melhores práticas de ambas, capazes de atender às peculiaridades de uma equipe composta por alunos em graduação. Para analisar a viabilidade do uso desse processo, está sendo desenvolvido um questionário avaliativo que será aplicado aos

membros das equipes que irão utilizá-la, comparando o nível de agilidade entre as metodologias XP, RUP, Scrum e a proposta através do uso dos Princípios Ágeis [Fowler, 2001], como também o nível de satisfação da equipe em utilizá-la.

4. Conclusão

Através da combinação das melhores práticas das metodologias XP, RUP e Scrum, e da observação de limitações encontradas na sua utilização em *software houses*, este trabalho propõe adaptar o processo de desenvolvimento de software em âmbito acadêmico, apresentando diretrizes para uma nova proposta de metodologia que busca o conhecimento de mercado para o desenvolvimento de sistemas em projetos acadêmicos dentro da Universidade.

Resultados preliminares apontam o uso majoritário de metodologias ágeis nas empresas regionais, demonstrando um alinhamento entre academia e mercado local. Entretanto, a academia ainda preocupa-se em distribuir o conteúdo igualmente entre as diversas metodologias, enquanto que no mercado local predomina o Scrum. Ao final da pesquisa, conforme os resultados sejam consolidados a partir do que foi observado e relatado, este conjunto de diretrizes servirá de embasamento teórico para a proposta de uma nova metodologia de desenvolvimento de software em âmbito acadêmico que incorpore práticas ágeis comumente utilizadas fora desse ambiente, mas com utilização de métodos de ensino e prática necessários para um bom aprendizado, facilitando a integração de egressos de cursos da área no mercado de trabalho.

Referências

Beck, Kent. (2004) “Extreme Programming Explained: embrace change”. 2st ed. Addison-Wesley.

Fowler, M. (2001) “The new methodology”, Wuhan University Journal of Natural Sciences. Vol. 6, n. 1-2, p. 12-24.

Garcia, Y. P. C., et. al. (2004) “easYProcess: Um Processo de Desenvolvimento para Uso no Ambiente Acadêmico”, In: XII Workshop de Educação em Informática - XXIV Congresso da Sociedade Brasileira de Computação, Salvador, BA.

Ludvig, D. e Reinert, J. D. (2007) “Estudo do uso de Metodologias Ágeis no Desenvolvimento de uma Aplicação de Governo Eletrônico”, Trabalho de Conclusão de Curso, UFSC, Florianópolis.

Pham, A. e Pham, P. (2011) “Scrum em Ação: gerenciamento e desenvolvimento ágil de projetos de software”. São Paulo: Novatec.

Rodrigues, N. N.; Estrela, N. V.A. (2012) “Simple Way: Ensino e Aprendizagem de Engenharia de Software Aplicada através de Ambiente e Projetos Reais”. In: VIII *Simpósio Brasileiro de Sistemas de Informação*, São Paulo, p. 722–733.

Sbrocco, J. H. T. C. Macedo, P. C. (2012) “Metodologias ágeis: Engenharia de software sob medida”. São Paulo: Erica.

Soares, M. S. (2004) “Metodologias Ágeis Extreme Programming e Scrum para o desenvolvimento de software”. Revista Eletrônica de SI, vol. 3, p. 8-13.

Tsui, F. F. (2013) “Fundamentos da Engenharia de Software”, Rio de Janeiro: LTC 2 ed.

Aplicação de Rede Bayesiana na identificação de fatores que causam a evasão no curso de Sistemas de Informação da UFRRJ.

Ramon Lameira[1], Robson Mariano da Silva[1]

[1] Departamento de Matemática/Sistemas de Informação – Instituto de Ciências Exatas – Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Caixa Postal 15.064 – 91.501-970 – Seropédica, RJ – Brasil

ramon_lameira@hotmail.com, robsonms@ufrrj.br

Abstract. *We propose a computational model based on the use of Bayesian networks in order to identify factors leading to evasion in the undergraduate course in the Information Systems – UFRRJ. The proposed model was capable of identify the non-correlation of origin as school dropout factor. The sensitivity obtained in the forecast of evasion in the validation set was 93.3%. The results show promising regarding on the use of bayesian network, in identification of factors and prediction of student evasion.*

Resumo. *Propomos um modelo computacional baseado na utilização de redes bayesianas, a fim de identificar fatores que levam a evasão no curso de graduação em Sistemas de Informação da UFRRJ. O modelo proposto foi capaz de identificar a não correlação da origem escolar como fator de evasão. A sensibilidade obtida na predição de evasão no conjunto de validação foi de 93,3%. Os resultados obtidos mostram-se promissores quanto ao uso de rede bayesiana, na identificação de fatores e predição da evasão discente.*

1. Introdução

A evasão escolar é, certamente, um dos problemas que afligem as instituições de ensino em geral gerando prejuízos sociais, econômicos, e acadêmicos a todos os envolvidos no processo educacional. Segundo Martinho [2012], é um problema complexo advindo da superposição de fatores sociais, econômicos, culturais e acadêmicos, além de variáveis demográficas e atributos individuais que influenciam na decisão do estudante permanecer ou abandonar o curso.

Diante deste quadro, nos últimos anos, no Brasil, foram realizados monitoramentos, estudos empíricos, pesquisas científicas e levantamento estatístico sobre a evasão no ensino superior, com vistas à adoção de medidas visando mitigar os altos índices de evasão. Dentre as contribuições e ações resultantes desses processos podemos relaciona-las sob três paradigmas; 1- contribuições teóricas, que visam compreender e identificar os fatores que contribuem e influenciam no processo de evasão; 2 – ações específicas de cada instituição de ensino superior (IES), estruturada em sua realidade e especificidade; e, 3 – ações governamentais, como o Programa de Reestruturação e Expansão das universidades Federais (REUNI), [MEC/SESu, 2007], estabelecido, com as Instituições Federais de ensino Superior (IFEs).

Apesar do esforço dos órgãos governamentais, os estudos sobre evasão existentes no Brasil, ainda são incipientes e tímidos, se comparados com os países desenvolvidos onde os estudos e pesquisas sistematizados são inúmeros [Portela, 2013].

Neste contexto, é imprescindível realizar estudos que possibilite a identificação precoce de estudantes vulneráveis à evasão, de sorte a possibilitar a aplicação de ações proativas no sentido de evitar o abandono.

Nos últimos anos, diversas metodologias baseada em inteligência computacional, foram desenvolvidas visando avaliar a evasão escolar. Dentre essas metodologias, pode ser citado o trabalho de Mustafá [2012] que propõe a aplicação de árvore de regressão e classificação para identificar a evasão discente, tendo como base inicial os dados da inscrição do estudante em um curso presencial. Os resultados obtidos permitiram concluir que os dados utilizados conferem um baixo nível de precisão na identificação da evasão.

Martinho *et al*., [2012] propuseram um sistema inteligente para a predição de grupos de risco de evasão discente nos cursos de tecnologia do Instituto Federal de Mato Grosso, utilizando rede neural ARTMAP-*Fuzzy*. Os resultados obtidos apresentaram índice de acerto do grupo evasivo de 97% e acerto global superior a 76%, valores estes que demonstram a confiabilidade, a acurácia e a eficiência do sistema.

O presente trabalho propõe um modelo computacional baseado em Redes Bayesianas para a identificação de fatores que causam a evasão discente no curso de graduação em Sistemas de Informação da UFRRJ.

Este artigo é estruturado da seguinte forma. Na Seção 1, a justificativa e o objetivo deste trabalho são descritos. A Seção 2 descreve os principais métodos e materiais utilizados ao longo do desenvolvimento deste trabalho. A Seção 3 mostra os resultados obtidos com a aplicação do modelo computacional em comparação com dados da literatura. A Seção 4, além de enfatizar as realizações que o trabalho conseguiu alcançar, também cita uma proposta de inserir novos parâmetros no modelo computacional.

2. Materiais e Métodos

O conjunto de dados constitui-se de 70% (n=105) dos alunos com matrícula no curso de Sistemas de Informação, obtidos junto a Pró-reitoria de Graduação da Universidade Federal Rural do Rio de Janeiro (UFRRJ). Sendo que desse universo ≈ 90% (n=90) foram selecionados para compor o conjunto de treinamento do modelo e o conjunto de validação com ≈10% (n=15) dos alunos com matrícula ativa ou não no referido curso escolhidos ao mero acaso. Com as seguintes variáveis: gênero (masculino ou feminino), origem escolar (pública ou privada), reprovação (sim ou não), distância (perto ou longe), para o parâmetro perto foi adotado uma distância inferior a 40 km da universidade e evasão (sim ou não).

O método proposto consiste na utilização de Redes Bayesianas, que é um modelo gráfico para relacionamentos probabilístico entre um conjunto de variáveis [Neapolitan, 2004]. Os sistemas estruturados em Rede Bayesiana são capazes de gerar automaticamente predições ou decisões mesmo na situação de inexistência de informações. Onde cada variável aleatória (VA) é representada por um nó da rede e cada nó de (VA) recebe conexões dos nós que têm influência direta sobre ele, além de possui uma tabela de probabilidade condicional, que quantifica a influência dos seus pais sobre ele. Cuja finalidade é a avaliar o impacto de cada variável na identificação de fatores que promovam a evasão e realizar a predição do grupo de estudantes, de

forma acurada e contínua, inferindo de maneira fidedigna. A estrutura da rede (Figura 1) foi estabelecida em função da disposição de dependências e independências entre as variáveis utilizadas. Para cada dependência entre os nós da rede, foi calculada a probabilidade *apriori* das variáveis utilizadas no modelo.

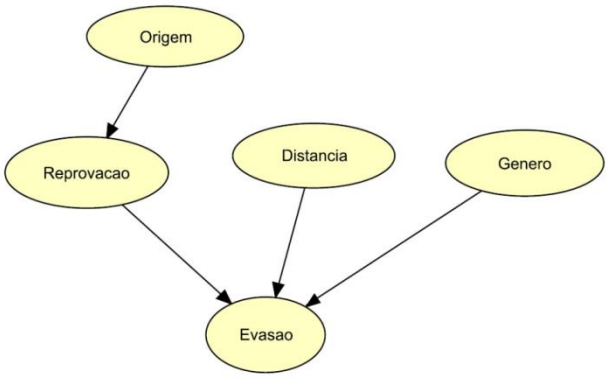


Figura 1. Fluxograma da Rede Bayesiana.

Para avaliar o desempenho do modelo neural bayesiano na predição de evasão no conjunto de teste, foram utilizados os índices de sensibilidade (verdadeiro positivo), especificidade (verdadeiro negativo). O modelo proposto foi implementado utilizando o pacote RHugin [Stefan e Achim, 2008] e o *software Hugin Expert* (www.hugin.com).

3. Resultados e Discussão

Neste trabalho foi desenvolvido uma rede bayesiana de sorte possibilitar a identificação de fatores que levam a evasão, bem como a sua predição . Os resultados obtidos pela inferência da rede implementada para o conjunto de dados em tela, em função das probabilidades propagadas, mostrou que não há evidência da contribuição da origem escolar na evasão dos discentes (Figura 2), fato este corroborado por [Nunes, 2013], que afirma que a origem não contribui para a evasão escolar. A taxa de evasão obtida foi de 72,3%, valor próximo ao descrito na literatura para a área de tecnologia da informação.

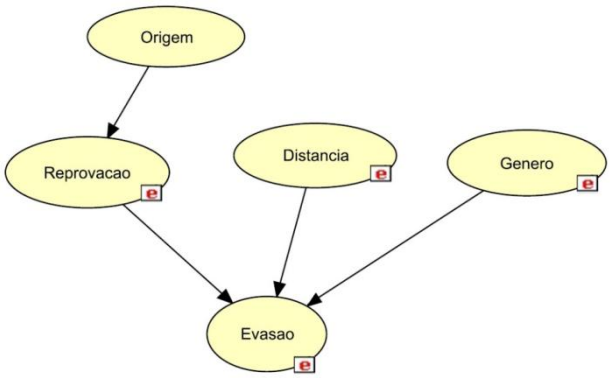


Figura 2. Dependências selecionadas pela rede bayesiana.

No que tange a variável reprovação o valor obtido pelo modelo foi de 89,1%, indicando uma forte correlação entre a reprovação e a evasão discente. Quanto à predição, o modelo computacional proposto obteve um erro de sensibilidade (verdadeiro positivo) para a evasão de 6,7%, e critério BIC -54,49% para o conjunto de validação (15 alunos).

4. Conclusão

A contribuição desse trabalho consistiu no desenvolvimento de um modelo computacional neural bayesiano na identificação e predição de evasão no curso de sistemas de Informação da UFRRJ.

Os resultados obtidos pelo modelo mostram-se promissores quanto à utilização da inferência bayesiana na identificação precoce dos fatores que levam a evasão no curso de Sistemas de Informação da UFRRJ. Além disso, o modelo foi capaz de identificar existência de correlação entre as variáveis e a evasão no referido curso.

Como proposta para trabalhos futuros pretende-se aplicar novas variáveis referentes a requisito didático/pedagógico, institucionais e sócio políticos/econômicos.

5. Referências

Martinho, V.R.C.(2012) "Sistema Inteligente para Predição do Grupo de Risco de Evasão Discente". Unesp, Faculdade de Engenharia. 58p.

Martinho, V. R. C.; Nunes, C.; Minussi, C. R. (2012) "Um sistema inteligente usando redes neurais artmap-fuzzy para predição do grupo de risco de evasão discente em cursos superiors presenciais". Simpósio Brasileiro de Automação Inteligente. Fortaleza, Brasil.

MEC – Ministério da Educação (2007). Diretrizes Gerais do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais.

Mustafa, M. N.; Chowdhury, L. and Kamal, S. (2012) "Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country". International Conference on Informatics, Electronics & Vision. Dhaka, Bangladesh: IEEE

Neapolitan, R.E. (2004)"Learning Bayesian Networks". Upper Saddle River: Pearson.

Nunes, R.C. (2013) "Panorama Geral da Evasão e Retençao no Ensino Superior no Brasil(IFES)". XXVI Encontro Nacional de Pró-reitores de Graduação.

Portela, S. (2013) "Evasão ou Retençao? Uma questão crucial à sustentabilidade das Instituições de Ensino Superior", ABMESeduca.com, São Paulo. (acessado em 10.08.2014).

Stefan, T., Achin, Z., (2008) "Collaborative software development usind R-Forge". Report 81, Departament of Statistics and Mathematics, Wirtschaftsuniversitat Wien, Research Report Series.

Uma abordagem para Sistemas de Informação Usando Realidade Aumentada e Dispositivos Móveis

Ana Beatriz Dreveck[1], **Alexandre da Costa Sena**[1,2], **Anselmo Antunes Montenegro**[3]

[1]Centro Universitário LaSalle (UNILASALLE)
Niterói – RJ – Brasil

[2]Universidade do Estado do Rio de Janeiro (UERJ)
Rio de Janeiro, RJ – Brasil

[3]Universidade Federal Fluminense (UFF)
Niterói, RJ – Brasil

asena@ime.uerj.br, anselmo@ic.uff.br

Abstract. *Mobile devices, especially smartphones, are part of the daily lives of people. In these devices, the applications are virtual environments that users can chat, take photos, play, among other activities. One way to make the interaction between user and application more real and interesting is using augmented reality, which is characterized by adding virtual supplements in real environments in real time. Thus, the user experience is more complete having not only the virtual environment in the mobile device, but also the influence of the real world. In this context, the aim of this paper is to propose an approach for the creation of information systems using mobile devices and augmented reality. Furthermore, to show the feasibility of the proposed approach, a case study application was developed to demonstrate how this interaction works and its benefits.*

Resumo. *Dispositivos móveis, especialmente os smartphones, fazem parte do dia a dia das pessoas. Nesses dispositivos, os aplicativos são ambientes virtuais que os usuários utilizam para conversar, tirar fotos, jogar, entre outras atividades. Uma maneira de tornar a interação entre usuário e aplicativo mais real e interessante é utilizar a realidade aumentada, que se caracteriza por adicionar suplementos virtuais em ambientes reais em tempo real. Assim, o usuário passa a ter não só o ambiente virtual do dispositivo móvel, como também, inserções virtuais no ambiente real em que estiver. Logo, o objetivo deste trabalho apresentar uma abordagem para criação de sistemas de informação utilizando dispositivos móveis e realidade aumentada. Além disso, para mostrar a viabilidade da abordagem proposta, foi desenvolvido um aplicativo como estudo de caso que demonstra como essa interação pode ser feita e os seus benefícios.*

1. Introdução

Os dispositivos móveis, em especial os *smartphones*, fazem parte do cotidiano das pessoas. No Brasil, cerca de 36% dos telefones móveis são *smartphones* e esse número não para de crescer conforme pesquisa feita em [Nielsen 2014], enquanto que no mundo são cerca de 56% [A. Hepburn 2013]. Mais importante, nesses aparelhos as pessoas gastam muito mais tempo utilizando aplicativos, cerca de 80% do tempo, do que propriamente falando com outras pessoas, conforme pesquisa divulgada em [A. Hepburn 2013].

Os aplicativos dos dispositivos móveis muitas vezes criam ambientes virtuais onde os usuários podem conversar, tirar fotos, jogar, desenhar, consultar mapas, dentre outras atividades. Uma maneira de tornar a interação entre usuário e aplicativo mais real e interessante é utilizar a realidade aumentada, que se caracteriza por adicionar suplementos virtuais em ambientes reais em tempo real [R. Tori 2006]. Desse modo, o usuário passaria a dispor não só o ambiente virtual do dispositivo móvel, como também, experimentar e usufruir das propriedades e caractersticas do ambiente real em que ele se encontra.

Nesse contexto, o objetivo deste trabalho é apresentar uma abordagem e uma arquitetura para criação de aplicações utilizando dispositivos móveis e realidade aumentada, descrevendo como essas aplicações podem ser usadas e suas principais vantagens. Assim, para mostrar a viabilidade da abordagem proposta, é desenvolvido um aplicativo como estudo de caso que utiliza dispositivos móveis tradicionais, como *smartphones* e *tablets*, e realidade aumentada.

Mais especificamente, é desenvolvido um jogo de estratégia baseado em cartas para ser jogado em ambientes móveis baseados na plataforma Android. As jogadas são realizadas nos dispositivos móveis, enquanto que as simulações sobre os efeitos das jogadas são mostradas usando realidade aumentada na tela de um computador.

2. Realidade Aumentada

A Realidade Aumentada (RA) é uma forma inovadora de se ver o mundo, que vem sendo estudada há aproximadamente 50 anos. No início, as pesquisas foram muito lentas devido à limitação tecnológica da época. Contudo, nos últimos anos, o avanço do hardware e da largura de banda impulsionaram esses estudos, possibilitando a evolução da RA ao ponto de sermos capazes de utilizá-la no nosso cotidiano [Bimber 2012].

Embora o uso da RA seja mais comum, ela faz parte de um contexto mais amplo, chamado de Realidade Misturada (RM). Esse termo é utilizado para definir a sobreposição de objetos virtuais tridimensionais com o ambiente físico real, onde o usuário pode visualizar esse ambiente gerado com o apoio de equipamentos tecnológicos. Isso acontece quando cenas reais e virtuais são misturadas em um ambiente só, sendo o ambiente real ou não. A RA acontece quando, em um ambiente de RM, temos o mundo real onde são colocados elementos criados por computador [R. Tori 2006].

Diferentemente da Realidade Virtual, a RA possibilita uma experiência mais real e natural com um aplicativo. Ao se utilizar acessórios portáteis, como o Google Glass ou um telefone celular, para capturar imagens em tempo real do ambiente físico, elementos virtuais são adicionados às imagens reais capturadas dando a impressão de que estes estão efetivamente no mundo real.

Até o momento, a RA pode ser obtida de quatro maneiras diferentes [C. Kirner 2005]: Sistema de visão ótica direta; Sistema de visão direta por vídeo; Sistema de visão por vídeo baseado em monitor; Sistema de visão ótica por projeção.

O **sistema de visão ótica direta** utiliza equipamentos, como óculos e capacetes, que não interferem na visão real do usuário. Para isso, eles possuem uma lente inclinada, onde são projetados os elementos virtuais devidamente ajustados ao mundo real. Por sua vez, o **sistema de visão direta por vídeo** utiliza dispositivos similares ao sistema de visão de ótica direta, porém, ao contrário deste, ele interfere na visão real do usuário.

Isso acontece porque ela é substituída por imagens de uma câmera de vídeo acoplada ao equipamento, mostradas ao usuário por meio de pequenos monitores inseridos nas lentes desses equipamentos. O óculus Rift é um exemplo de equipamento voltado para construção de sistemas de visão direta por vídeo.

Diferentemente dos outros, o **sistema de visão por vídeo baseado em monitor** utiliza dispositivos que não precisam ser equipados para se conseguir a RA. Esse sistema usa uma câmera de vídeo junto a um computador com monitor para capturar a imagem real e processá-la. Após o processamento da imagem, são inseridos objetos tridimensionais nela e depois ela é apresentada ao usuário por meio do monitor. Nesse sistema, o posicionamento da câmera normalmente é fixo, o que impede o deslocamento do usuário [C. Kirner 2005].

Por último, o **sistema de visão ótica por projeção** também não necessita de dispositivos equipáveis. A ideia é bem simples, onde imagens virtuais são projetadas no ambiente real, não sendo necessário nenhum dispositivo tecnológico para visualização dessas imagens. Para se utilizar esse sistema, pode ser necessário uma superfície de projeção, o que limita muito o seu uso [C. Kirner 2005].

As aplicações em RA que estão sendo desenvolvidas abrangem as mais diversas áreas, como saúde, entretenimento e educação [Bimber 2012]. A Disney vem apostando nessa tecnologia, principalmente no sistema de visão ótica por projeção, para dar mais vida e magia aos seus parques temáticos [Mine 2012]. Na área da medicina, estão sendo desenvolvidas aplicações em RA para ajudar ainda mais nas cirurgias [Nassir 2012].

3. Uma Abordagem para Criação de Sistemas de Informação com Dispositivos Móveis e Realidade Aumentada

O objetivo deste projeto é apresentar uma abordagem para criação de aplicações que utilizem realidade aumentada e dispositivos móveis. Para isto, é utilizado o sistema de visão por vídeo baseado em monitor para apresentar a realidade aumentada, por meio de um computador com câmera, como é apresentado na Figura 1.

Nesta abordagem, o computador funciona como servidor da aplicação e os dispositivos móveis seus clientes, de maneira que só há comunicação entre o servidor e os dispositivos móveis, onde essa comunicação pode ser feita por Bluetooth ou Wi-Fi. É importante notar que essa proposta de interação não fica limitada a um único usuário, podendo ser utilizada por diversos usuários ao mesmo tempo. Por exemplo, a Figura 1 mostra como seria a interação com dois usuários.

Na abordagem proposta, o(s) usuários interage(m) com o aplicativo através do dispositivo móvel e a realidade aumentada é visualizada através de um monitor. Enquanto que as informações principais são visualizadas no próprio dispositivo móvel, o monitor é utilizado para apresentar informações complementares através da realidade aumentada, que ajudam enriquecer a aplicação e, consequentemente, ampliar e melhorar as informações apresentadas.

Esta forma de interação onde o usuário utiliza dispositivos móveis em conjunto com realidade aumentada pode trazer benefícios para diversas áreas, como por exemplo, educação e entretenimento. Uma aula poderia ficar muito mais dinâmica e interessante se os alunos pudessem interagir através de um dispositivo móvel e visualizar parte do

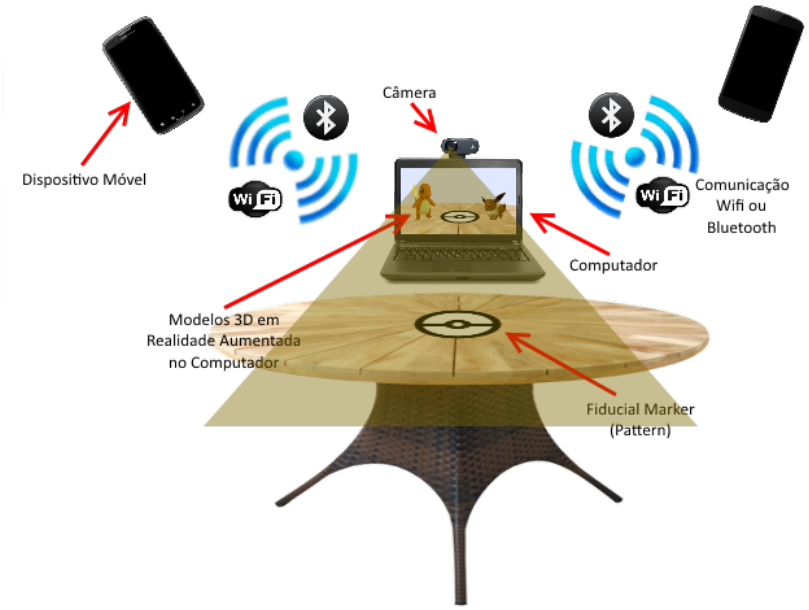


Figura 1. Aplicação Utilizando Dispositivos Móveis e Realidade Aumentada

conteúdo através da realidade aumentada. Por exemplo, em uma aula de alfabetização, os alunos teriam uma lista de palavras, onde teriam que indicar a palavra daquela lista que inicia com determinada letra usando um dispositivo móvel. Após a indicação de todos os alunos, um modelo 3D contendo uma imagem do significado da palavra correta apareceria para todos em realidade aumentada.

É importante ressaltar que, apesar da abordagem proposta utilizar o sistema de visão por vídeo baseado em monitor, uma forma ainda mais interessante e inovadora seria utilizar o sistema de visão ótica direta com o auxílio de um dispositivo ótico como o Google Glass. Nessa nova abordagem, a realidade aumentada seria visualizada por cada pessoa através de seu dispositivo ótico. A única razão para não utilizar essa abordagem neste trabalho foi a ausência de um exemplar do dispositivo ótico para validar a proposta.

A abordagem proposta permite a criação de ferramentas que possibilitam o usuário aprender através da exploração e descoberta, conforme o pensamento Construcionista de Papert [I. Harel 1991], ao invés de ferramentas onde o aluno simplesmente consulta uma informação, memoriza e repete.

4. Estudo de Caso - PokARTCG

A viabilidade da abordagem proposta para criação sistemas de informação utilizando dispositivos móveis e realidade aumentada é apresentada através da implementação de um jogo de cartas colecionáveis, chamado TCG (*Trading Card Game*). Existem diversos tipos de TCG, onde cada um tem suas próprias regras e elementos. Neste projeto é implementado o o jogo de cartas Pokémon Estampas Ilustradas, mais conhecido como Pokémon TCG, que neste trabalho foi chamado de PokARTCG.

4.1. Arquitetura do Sistema

Como o estilo de jogo padrão do Pokémon TCG necessita de dois jogadores, a arquitetura do sistema é composta de dois clientes e um servidor central, conforme apresentado na Figura 2. Os clientes só se comunicam com o servidor e ele é responsável por repassar as mensagens recebidas ao outro cliente, caso seja necessário. Nesse trabalho, a integração cliente-servidor é feita por meio de uma conexão Wi-Fi. Porém, não há nenhum impedimento para o uso de Bluetooth.

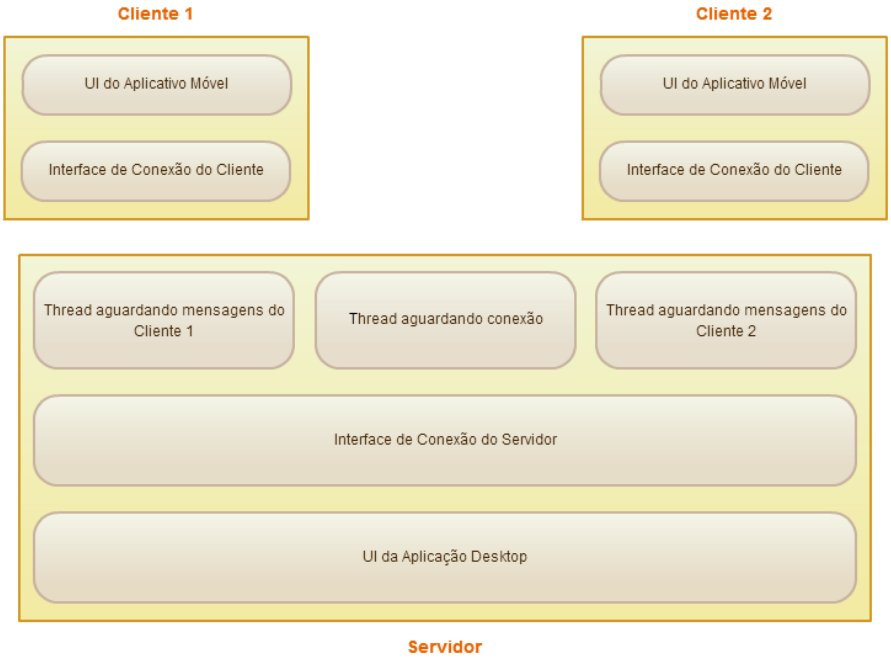


Figura 2. Visão Geral da Arquitetura do Sistema

Conforme apresentado na Figura 2, os clientes são basicamente compostos de uma interface com o usuário e uma interface de conexão. Por sua vez, o servidor possui um conjunto de *threads* esperando as conexões com os clientes, além de uma interface com o usuário e uma interface de conexão.

O servidor foi desenvolvido para ambiente Windows e é responsável por processar e repassar todas as mensagens recebidas. Ele fica ativo, exibindo em tempo real a imagem capturada pela câmera conectada a ele, esperando por duas conexões com dispositivos móveis clientes que possuam o aplicativo PokARTCG. Os clientes, por sua vez, são aplicativos para dispositivos móveis desenvolvidos para o sistema operacional Android, que implementam virtualmente o jogo Pokémon TCG.

Após os dois dispositivos se conectarem com o servidor, ele fica aguardando uma mensagem para exibir a realidade aumentada. Ao receber a mensagem, o servidor exibirá modelos 3D animados de pokémons lutando no campo de batalha. A posição de cada modelo é calculada baseada no *fiducial marker*, que deve ser posicionado no meio do campo de batalha, capturado pela câmera.

O aplicativo desenvolvido para os dispositivos móveis foi programado utilizando a linguagem Java para Android, com o auxílio do plugin ADT (Android Developer Tools) Bundle fornecido pela Google [ADT Bundle 2014]. O ADT Bundle é um conjunto

de ferramentas para desenvolvimento de aplicativos para Android, contendo, entre outras ferramentas, o Android SDK e o ambiente de desenvolvimento Eclipse, já com o plugin ADT. Para facilitar a criação do jogo, foi utilizada a *engine* para jogos AndEngine GLES2 [AndEngine 2014], criada especialmente para desenvolvimento de jogos para plataformas Android.

O aplicativo servidor, que pode ser executado em uma máquina *desktop* ou *notebook*, foi desenvolvido utilizando a engine Unity 3D [Unity 2014], usada principalmente para o desenvolvimento de jogos 3D. As linguagens dos scripts gerados para Unity 3D, são feitas, principalmente, em C# e JavaScript. Para criação da realidade aumentada, foi integrada ao projeto a biblioteca NyARToolkit [NyARToolkit 2014] para Unity, que é derivada do ARToolKit (Augmented Reality Tool Kit). O ARToolKit é um conjunto de bibliotecas criado usando técnicas para captura de imagens das fontes disponíveis no computador e para rastrear os respectivos marcadores nas imagens capturadas. Além disso, ao achar o marcador indicado, essas bibliotecas incorporam modelos gerados por computador as imagens, em posições baseadas nos marcadores. Por último, para auxiliar a utilização do NyARToolkit, foi usada a biblioteca Simplify!, que contêm códigos pré-fabricados, permitindo um desenvolvimento de RA simples e rápido.

As Figuras 3 e 4 são imagens do aplicativo cliente, que roda no dispositivo móvel, e do servidor, que roda em um computador comum, respectivamente. A Figura 3 representa o campo de batalha do jogo onde acontecem todas as ações. A partir do momento que um jogador realiza um movimento e finaliza sua jogada, uma simulação em realidade aumentada do efeito da jogada é apresentado no servidor. Por exemplo, a Figura 4 mostra a realidade aumentada de dois pokémons no campo de batalha.

Figura 3. Campo de Batalha do aplicativo para dispositivo móvel

5. Trabalhos Relacionados

O trabalho em [Henrysson 2007] explica como a realidade aumentada vem sendo estudada há décadas, atraindo cada vez mais a atenção das pessoas, sem conseguir um grande número de usuários. Para alcançar um público maior, Henrysson propõe que a RA seja

Figura 4. Campo de Batalha em Realidade Aumentada na Tela do Servidor

vista através de dispositivos móveis e apresenta algumas soluções para tornar possível essa proposta, devido à capacidade limitada de processamento desses dispositivos.

Por sua vez, o trabalho apresentado em [M. A. Galvão 2012] mostra como a realidade aumentada pode enriquecer a leitura e facilitar o entendimento de livros infantis. Eles desenvolveram dois estudos de caso onde dispositivos móveis foram utilizados para exibir as figuras em 3D e permitir uma interação em tempo real com os livros.

A Sony criou o jogo The Eye of Judgment [Sony 2014], um jogo de cartas colecionáveis, feito exclusivamente para a plataforma Playstation 3 (PS3). O jogo inova na área de Trading Card Games (TCG), pois utiliza uma câmera exclusiva do PS3 para saber as cartas jogadas no campo de batalha. Com isso, todas as jogadas feitas são exibidas em 3D na televisão, dando a impressão de que cada ser místico jogado é real.

O ARBattle [T. Malheiros 2012] também é um jogo de cartas em realidade aumentada. Nele, a tela mostrada pela câmera é dividida entre dois jogadores. Assim, cada jogador pode usar uma carta com um marcador desenhado no espaço designado. Quando o programa reconhece o marcador desenhado na carta, o personagem virtual relativo à carta aparece em cima dela em realidade aumentada.

A Google também está investindo em realidade aumentada. A empresa desenvolveu o Ingress [Google 2014], um aplicativo para Android que utiliza a realidade aumentada para transformar o mundo real em um cenário de um jogo estratégico, onde o usuário precisa percorrer locais reais para realizar novas descobertas e avançar com o conteúdo.

Por fim, o colAR [Puteko 2014] é um aplicativo para dispositivos móveis voltado ao público infantil. Nele, as crianças imprimem um desenho para colorir, disponibilizado pela Puteko. Após colorirem a imagem, elas utilizam o aplicativo para visualizar o modelo virtual do desenho colorido em realidade aumentada, que aparece acima da imagem na tela do dispositivo.

O trabalho apresentado neste artigo é diferente de todos os outros, pois propõe uma abordagem e uma arquitetura para criação de sistemas que combinam a interação através dos dispositivos móveis e a visualização através da realidade aumentada.

6. Conclusão

Os dispositivos móveis estão cada vez mais presentes no cotidiano das pessoas, que utilizam seus aplicativos para as mais diferentes tarefas. Por sua vez, o avanço tecnológico

ajudou a evolução da realidade aumentada, ao ponto de ser possível utilizá-la até mesmo nesses aparelhos.

O principal objetivo deste trabalho foi propor uma abordagem e uma arquitetura para criação de sistemas que combinem o uso de dispositivos móveis com a realidade aumentada, tornanda a aplicação mais real e interessante. É esperado que este projeto beneficie diversas áreas, como educação, entretenimento, entre outras.

Referências

A. Hepburn (último acesso 30/12/2013). Infographic: 2013 Mobile Growth Statistics. http://www.digitalbuzzblog.com/infographic-2013-mobile-growth-statistics.

ADT Bundle (último acesso 01/10/2014). ADT Bundle: Google. http://developer.android.com/sdk/index.html.

AndEngine (último acesso 1/10/2014). AndEngine. http://www.andengine.org.

Bimber, O. (2012). What's real about augmented reality. *IEEE Computer*, 45(7):24–25.

C. Kirner, E. Z. (2005). Aplicações educacionais em ambientes colaborativos com realidade aumentada. In *Simpósio Brasileiro de Informática na Educação*.

Google (último acesso 04/10/2014). Ingress. https://www.ingress.com/.

Henrysson, A. (2007). *Bringing Augmented Reality to Mobile Phones*. PhD thesis, Universidade de Linkping, Norrkping.

I. Harel, S. p. (1991). *Constructionism*. Ablex Publishing.

M. A. Galvão, E. R. Z. (2012). Aplicaões móveis com realidade aumentada para potencializar livros. *RENOTE- Revista Novas Tecnologias na Educação*, 10(1).

Mine, M. (2012). Projection-based augmented reality in disney theme parks. *IEEE Computer*, 45(7):32–40.

Nassir, N. (2012). First deployments of augmented reality in operating rooms. *IEEE Computer*, 45(7):48–55.

Nielsen (último acesso 30/09/2014). O consumidor Móvel: Um panorama global. http://www.nielsen.com/br/pt/insights/reports/2013/o-consumidor-movel.html.

NyARToolkit (último acesso 05/10/2014). Nyartoolkit project. http://nyatla.jp/nyartoolkit/wp/?page_id=198.

Puteko (último acesso 04/10/2014). coLAR. http://colarapp.com/.

R. Tori, Cláudio Kirner, R. S. (2006). *Fundamentos e tecnologia de realidade virtual e aumentada*. SBC.

Sony (último acesso 04/10/2014). The Eye of Judgment. http://www.playstation.com/en-us/games/the-eye-of-judgment-ps3/.

T. Malheiros, L. R. (2012). A utilização da realidade aumentada em jogos de cartas colecionáveis. In *Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames2012)*.

Unity (último acesso 05/10/2014). Unity technologies. http://unity3d.com/pt.

Manutenção Automática dos Artefatos de Modelos de Processos de Negócio e dos Textos Correspondentes: Métodos e Aplicações

Raphael A. Rodrigues[1], **Leonardo G. Azevedo**[1,2], **Kate Revoredo**[1], **Henrik Leopold**[3]

[1]Programa de Pós-Graduação em Informática(PPGI)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Av. Pasteur, 456 – Urca – Rio de Janeiro – RJ – Brasil

[2]IBM Research
Av. Pasteur 146 & 138, Botafogo
Rio de Janeiro, RJ 22290-240 - Brasil

[3]WU Vienna, Welthandelsplatz 1, 1020 Vienna, Austria

{raphael.rodrigues, azevedo, katerevoredo}@uniriotec.br,

LGA@br.ibm.com, henrik.leopold@wu.ac.at

Abstract. *Business analysts are responsible for the validation of Business Process Models. They can easily read the model but do not have the necessary knowledge about the business. In the other hand, domain specialists know the business but do not have the necessary modeling skills. It is easier for them to read a natural language text. For this reason, both model and text, are necessary artifacts for a natural communication between business specialits and analysts. The manual management of both resources may result in inconsistencies, due to unilateral modifications. This research propose a methodology for text generation from process models and vice-versa. The result will be evaluated in real scenarios, through companies and case study.*

Resumo. *A validação de modelos de processos de negócio normalmente é executada por analistas de negócio. Estes podem ler facilmente o modelo, porém não detêm conhecimento aprofundado do negócio. Em contrapartida, especialistas de domínio conhecem o domínio do negócio, porém não detêm conhecimento sobre modelagem. Para estes, é mais fácil ler um texto em linguagem natural. Por este motivo, tanto o modelo de processo, quanto o texto descritivo, são artefatos necessários para permitir uma comunicação natural entre os especialistas e os analistas de negócio. Esta pesquisa propõe uma metodologia para geração de textos a partir de modelos e vice-versa. O resultado será avaliado em cenários reais, através de empresas e estudos de caso.*

1. Introdução

Organizações normalmente possuem uma sequência estruturada de atividades, de modo a completar uma determinada tarefa. Estas atividades precisam ser executadas em uma ordem específica, além de possuírem determinados papeis responsáveis pela sua execução.

É importante manter a documentação da estrutura de processos relacionadas com as atividades de uma organização [Kettinger et al. 1997]. Para grandes organizações, este requisito torna-se ainda mais importante [Davies et al. 2006], devido a existência de projetos grandes e complexos, exigindo uma abordagem mais detalhada para execução de suas atividades [Kautz et al. 2004]. Basicamente, Kautz et al. (2004) afirma, através de estudos de casos feitos numa grande companhia de software em Danish (Dinamarca), que o uso de uma metodologia formal aumenta a confiança necessária e a comunicação entre os empregados para execução de suas tarefas. Em grandes organizações é complexo entender a tarefa a ser executada sem antes ter uma visão do todo. Esta visão ampla, pode ser obtida através de metodologias como a modelagem conceitual. Por exemplo, com respeito a uma determinada atividade, se existe algum tipo de documentação sobre o processo (e.g, modelos de processos), torna-se mais fácil entender como a atividade em questão colabora para a execução do processo como um todo, além de detalhar os principais envolvidos no processo [Davies et al. 2006].

Diversas organizações utilizam notações para descrever seus processos de negócio, como a BPMN [Ko et al. 2009]. A notação BPMN foi desenvolvida e padronizada pelo *Object Management Group* [OMG 2011]. Apesar de notações como BPMN serem úteis em diferentes cenários, ainda é um grande desafio para muitos empregados entender a semântica de um modelo de processo. Se o leitor não está familiarizado com o grande conjunto de conceitos e elementos (por exemplo, *gateways*, eventos ou atores), partes do processo poderão não ser entendidas ou até mal interpretadas pelo leitor. Inclusive, diversos estudos sobre o entendimento de modelos de processos evidenciaram quão complexo a compreensão de modelos poderia ser, até mesmo para pessoas que estão familiarizadas com modelagem de processo [Mendling et al. 2012], [Figl and Laue 2011], [Reijers and Mendling 2011]. O treinamento de empregados no entendimento de modelos de processos está associado com uma considerável quantidade de tempo e dinheiro e dificilmente poderá ser considerado como uma opção viável para todos os empregados de uma organização.

Como uma solução para este problema, muitas organizações mantêm ambos artefatos, modelos e textos. Isto no entanto, significa que as empresas encontram esforços redundantes para atualizar ambos artefatos. O cenário se torna ainda mais crítico, considerando que inconsistências podem ocorrer devido a atualização unilateral dos artefatos, em outras palavras, somente um dos artefatos é modificado ou atualizado. Aparentemente, a manutenção em paralelo de ambos artefatos, modelos de processos e instruções de trabalho através de textos, não representa uma solução prática.

Para solucionar este problema de inconsistência, permitindo que as organizações evitem trabalho redundante, diversas técnicas de verbalização foram propostas. Estas técnicas permitem um mapeamento direto de modelos conceituais para linguagem natural [Nijssen and Halpin 1989]. Em pesquisas anteriores, nós elaboramos as primeiras propostas para a transformação de modelos de processos para textos em linguagem natural [Leopold et al. 2012], [Leopold et al. 2014], [Rodrigues 2013]. Porém, enquanto que esta técnica permite a geração de texto a partir de modelos de processos, não oferece a oportunidade de refletir alterações no texto para o modelo de processo utilizado como entrada. Apesar da existência de algumas técnicas que lidam com o mapeamento de modelos de processos para linguagem natural, até onde sabemos, não existem técnicas que

possibilitem a transformação bilateral entre os artefatos, i.e, refletindo as alterações de um modelo de processo para o texto e vice versa (*round-trip*). Considerando esta situação na prática, tal técnica está associada com um conjunto de benefícios consideráveis:

- Se usuários do negócio não estão familiarizados com modelos de processos, textos em linguagem natural podem ser automaticamente gerados, alterados, e transformados novamente para modelos de processos.
- Textos que representam modelos são ativos importantes para a organização, uma vez que discussões baseadas em textos tendem a ser mais produtivas [Leão et al. 2013].
- Linguagem natural permite unificar a visão dos processos da organização, considerando que todos as áreas da empresa conseguem entender textos em linguagem natural.
- Se um dos artefatos for alterado, as alterações podem ser refletidas automaticamente para sua respectiva contra parte.
- Não é mais necessário manter ambos os artefatos. Uma organização pode simplesmente mater um repositório de modelos de processos. Os textos em linguagem natural podem ser gerados automaticamente a qualquer momento e estarão sempre consistentes com os respectivos modelos de processos.

Considerando os benefícios associados com a abordagem supracitada, este projeto tem como principal objetivo a definição de uma técnica que seja capaz de gerar texto a partir de modelos de processos, além de permitir a alteração automática do modelo a partir de modificações no texto. Visando prover a maior flexibilidade possível, nós temos como meta o desenvolvimento da técnica, de modo que ela seja adaptável para diversos idiomas, incluindo inglês, português e alemão. Diversos algoritmos serão desenvolvidos, permitindo a execução automática da abordagem *round-trip*.

A principal meta a ser alcançada é a criação de uma abordagem que seja capaz de interpretar modelos de processos de negócio, escritos em diferentes idiomas pertencentes a ramificação Romana e Germânica da família de linguagens Indo-Europeias, como por exemplo: português, alemão, inglês e espanhol. Para tal fim, deverão ser analisadas as diferentes anotações para a modelagem de processo, visando a identificação de um modelo canônico, para a representação de modelos de processos descritos em diferentes notações. Além do estudo linguístico dos idiomas considerados nesta proposta, um experimento será conduzido a fim de identificar: (1) se usuários que não possuem experiência com modelagem de processos, preferem ter acesso a um texto em linguagem natural que descreve o processo e (2) qual é a influência que o conhecimento em modelagem de processos exerce sobre o entendimento de um processo organizacional.

Como metas secundárias, visa-se estender a aplicabilidade de extração de informações linguísticas a partir modelos de processos, para possibilitar tarefas como:

- Validação de regras de nomenclaturas e design de processos
- Sumarização automática de modelos de processos
- Elaboração de um modelo canônico que permita mapear modelos escritos em diferentes notações para sua respectiva forma canônica.

Por último, este projeto de pesquisa está relacionado a área de Computação e Tecnologia da Informação. Além disso, está diretamente relacionado ao item “Gestão da

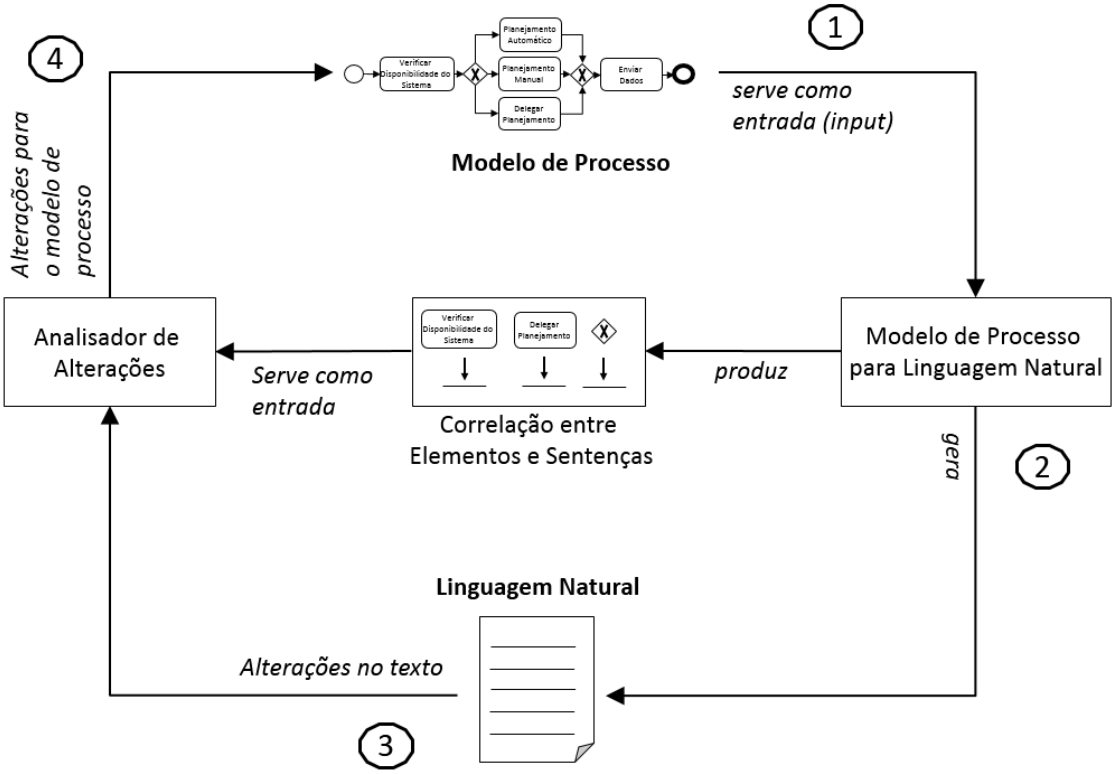


Figura 1. Exemplo da arquitetura para a solução proposta.

informação em grandes volumes de dados multimídia distribuídos" proposto como desafio pela Sociedade Brasileira de Computação como "Grandes Desafios da Computação Brasil: 2006-2016".

2. Proposta de Projeto de Pesquisa

A proposta deste projeto de pesquisa é manter automaticamente os artefatos de processo de negócio atualizados, tanto os textos descritivos do processo quanto os modelos gráficos que o representam.

2.1. Visão Geral da Proposta

O objetivo final do projeto é definir e desenvolver uma técnica bidirecional (*round-trip*) para a transformação automática de modelos de processos em linguagem natural e vice versa.

A figura 1 detalha a visão geral de nossa arquitetura proposta para a solução bidirecional. A ideia principal de nossa arquitetura é manter um conjunto de correlações entre elementos do modelo de processo e do texto gerado. Baseando-se na construção de nossa técnica de geração anterior [Leopold et al. 2012] (passo 1 e 2), tal correlação pode ser produzida automaticamente. Se um usuário aplicar alterações no texto em linguagem natural (passo 3), estas alterações serão analisadas pelo `Analisador de Alterações`. Este analisador será o núcleo de nossa técnica bidirecional. Ao referenciar as ligações entre os elementos do modelo e sentenças em linguagem natural, ele detecta quais foram os elementos do modelo afetados pela alteração textual e como estas alterações impactam

no modelo original. No passo subsequente (passo 4), as alterações no texto são refletidas para o modelo de processo. Cabe ressaltar que o acionamento do Analisador de Alterações está vinculado a detecção de alterações no texto. Em outras palavras, este componente somente será acionado caso haja alterações no texto.

Além do design dos componentes principais, nós estendemos os objetivos considerando três dimensões importantes:

1. **Multilinguistico**: Grande parte das organizações internacionais consideram o Inglês como sua linguagem de negócio padrão. Porém, de modo a evitar desentendimentos, muitas empresas ainda criam modelos de processos em sua língua nativa. Logo, nós propomos uma abordagem multilinguística. Atualmente, todos os trabalhos sobre geração de linguagem natural a partir de modelos de processos, e geração de modelos a partir de textos em linguagem natural, estão limitados para um único idioma [Meziane et al. 2008], [Lavoie et al. 1996], [Dalianis 1992].
2. **Independência da Linguagem para Modelagem**: Na indústria diversas linguagens de modelagem são utilizadas. Frequentemente, encontramos exemplos de modelos elaborados em BPMN, EPCs e UML. Reconhecendo esta característica, é nosso objetivo definir uma abordagem bidirecional que não esteja limitada a apenas uma única notação específica, e.g, BPMN. As soluções existentes focam em uma única linguagem de modelagem (e.g, [Meziane et al. 2008], [Lavoie et al. 1996]). Nós almejamos definir um formato de modelagem que seja independente de linguagem, que possa ser mapeado para qualquer notação utilizada para elaboração de modelos de processos. Como resultado, nossa abordagem bidirecional poderá ser empregada para diferentes linguagens de modelagem.
3. **Solução *User-friendly***: Apesar de ser um projeto de pesquisa, acreditamos que é importante definir uma técnica que possa ser eficientemente utilizada na prática. Desta forma, planejamos integrar nossa abordagem em um editor que esteja disponível ao público, como por exemplo, o editor online Signavio (`www.signavio.com`).

2.2. Objetivos da Proposta

A figura 1 ilustra o principal objetivo desta proposta de projeto de pesquisa. O objetivo contempla a definição e desenvolvimento de métodos e ferramentas para a manutenção automática de artefatos de processo de negócio. As técnicas desenvolvidas irão permitir gerar textos a partir de modelos, detectar alterações nos textos gerados e refletir estas alterações atualizando o modelo original. Em outras palavras, definiremos técnicas para a produção de textos a partir de modelos de processos de negócio, permitindo que os modelos sejam refinados através de alterações nos textos gerados. Em particular os modelos confeccionados utilizando-se a notação BPMN serão considerados. Uma primeira abordagem para a geração de texto a partir de modelos, flexível para diferentes idiomas, foi proposta por [Rodrigues 2013]. O objetivo agora é criar uma abordagem capaz de manter os artefatos automaticamente atualizados, refletindo as mudanças detectadas no texto para o modelo de processo que foi utilizado como base para sua confecção.

Adicionalmente, acredita-se que diversos objetivos secundários poderão ser atingidos, através da criação de algoritmos e aplicações de métodos para a extração de informações linguísticas a partir de modelos de processos, bem como o uso de proces-

samento de linguagem natural para atualizar um modelo baseando-se em alterações no texto.

Após construídas, as ferramentas serão disponibilizadas na forma de um framework que poderá ser integrado como um plug-in para ferramentas de modelagem de processos (e.g, editor online Signavio). Acredita-se que este plug-in será de grande serventia tanto para a indústria, quanto para a academia.

2.3. Metodologia

A metodologia a ser aplicada será baseada na proposta de métodos e aplicações destes, além da especificação de experimentos para avaliação dos mesmos. Detalhando esta abordagem temos:

1. Levantamento e estudo aprofundado dos trabalhos realizados na área, assim como dos conceitos, casos, estudos e técnicas relacionados aos seus temas de origem:
    - Levantamento e pesquisa sobre o estado da arte na área de GLN, em particular, considerando modelos de processos;
    - Levantamento e pesquisa sobre o estado da arte na área de Modelagem de Processos;
    - Levantamento e pesquisa sobre o estado da arte na área de Revisão de Teorias;
    - Levantamento e pesquisa sobre o estado da arte na área de Aprendizado de Modelos a partir de textos e Refinamento de Modelos;
2. Monitoramento de novos trabalhos e propostas publicados em congressos, periódicos e simpósios que estejam no escopo do presente projeto.
3. Desenvolvimento de métodos abordando as propostas do presente projeto.
4. Avaliação dos métodos desenvolvidos.
    - Inicialmente com dados artificiais, o que permitirá um experimento controlado.
    - Depois utilizando dados reais, de cenários a serem definidos.
5. Criação de novas parcerias e colaborações nacionais e internacionais, através de co-autorias, co-orientações e troca de conhecimentos adquiridos.
6. Avaliação e publicação dos resultados parciais e finais.

O trabalho produzirá resultados em duas frentes. Por um lado, serão desenvolvidos algoritmos para geração de textos a partir de modelos de processos, para a detecção de alterações feitas no texto gerado e para permitir a atualização automática do modelo original a partir das alterações detectadas. Por outro lado, estes algoritmos serão empacotados em uma ferramenta flexível, capaz de produzir textos a partir de modelos de processos escritos em diferentes idiomas, permitindo a atualização automática do modelo original a partir de alterações no texto. Além disso, a ferramenta será testada em casos reais e através de um estudo de caso, conduzido tanto com empresas quanto com universidades.

3. Conclusão

A execução deste projeto de pesquisa é um empreendimento interdisciplinar. Técnicas e conhecimento sobre diferentes campos de pesquisa, como linguística, processamento de linguagem natural, geração de linguagem natural, modelagem de processo, e engenharia de requisitos precisam ser combinados. Em particular, as seguintes contribuições serão consideradas:

- Mapear a influência que a experiência do usuário com modelagem de processos pode exercer sobre o entendimento e preferência pela representação de processos, através de textos ou modelos.
- Definir o ponto de convergência entre o entendimento de modelos e de textos. Em outras palavras, definir o grau de experiência necessário para que o usuário entenda melhor um modelo em detrimento do texto.
- Apresentar a correlação entre elementos do modelo de processo e elementos da linguagem natural.
- Técnicas para a análise e classificação automática de alterações textuais.
- Técnicas para refletir as alterações textuais para os modelos de processos.
- Técnicas para a classificação das alterações nos textos em linguagem natural.
- Desenvolvimento de uma abordagem bidirecional, eficiente e efetiva , integrada entre modelos e textos.

Está incluso neste projeto de pesquisa, um experimento com pessoas de diferentes organizações internacionais (tanto na indústria quanto na academia), com o intuito de verificar qual representação de processo, e.g modelos ou textos, é mais fácil de ser entendida pelo usuário, de acordo com a experiência em modelagem de processos.

Por fim, este projeto de pesquisa tem foco em algoritmos de geração de texto a partir de modelos de processos, permitindo também a atualização de modelos a partir de alterações nos textos gerados. Pretende-se, através dos algoritmos desenvolvidos, a criação de uma ferramenta capaz de gerar um texto em linguagem natural a partir de um modelo de processo elaborado em qualquer idioma, pertencente a ramificação Romana e Germânica da família de linguagens Indo-Europeias. Além da geração automática para qualquer idioma, a ferramenta será capaz de refletir as alterações do texto para o modelo de processo original, mantendo ambos artefatos sincronizados automaticamente.

Referências

Dalianis, H. (1992). A method for validating a conceptual model by natural language discourse generation. In *Advanced Information Systems Engineering*, pages 425–444. Springer.

Davies, I., Green, P., Rosemann, M., Indulska, M., and Gallo, S. (2006). How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3):358–380.

Figl, K. and Laue, R. (2011). Cognitive complexity in business process modeling. In *Advanced Information Systems Engineering*, pages 452–466. Springer.

Kautz, K., Hansen, B., and Jacobsen, D. (2004). The utilization ofínformation systems development methodologies in practice.

Kettinger, W. J., Teng, J. T., and Guha, S. (1997). Business process change: a study of methodologies, techniques, and tools. *MIS quarterly*, pages 55–80.

Ko, R. K., Lee, S. S., and Lee, E. W. (2009). Business process management (bpm) standards: a survey. *Business Process Management Journal*, 15(5):744–791.

Lavoie, B., Rambow, O., and Reiter, E. (1996). The modelexplainer. In *Demonstration presented at the Eighth International Workshop on Natural Language Generation, Herstmonceux, Sussex*.

Leão, F., Revoredo, K., and Baião, F. (2013). Learning well-founded ontologies through word sense disambiguation. In *Intelligent Systems (BRACIS), 2013 Brazilian Conference on*, pages 195–200. BRACIS.

Leopold, H., Mendling, J., and Polyvyanyy, A. (2012). Generating natural language texts from business process models. In *Advanced Information Systems Engineering*, pages 64–79. Springer.

Leopold, H., Mendling, J., and Polyvyanyy, A. (2014). Supporting process model validation through natural language generation. In *Transations on Software Engineering 40(8)*, pages 818–840. IEEE.

Mendling, J., Strembeck, M., and Recker, J. (2012). Factors of process model comprehension findings from a series of experiments. *Decision Support Systems*, 53(1):195–206.

Meziane, F., Athanasakis, N., and Ananiadou, S. (2008). Generating natural language specifications from uml class diagrams. *Requirements Engineering*, 13(1):1–18.

Nijssen, G. M. and Halpin, T. A. (1989). *Conceptual Schema and Relational Database Design: a fact oriented approach*. Prentice-Hall, Inc.

OMG (2011). Business process model and notation (bpmn) version 2.0. `http://www.bpmn.org/`.

Reijers, H. A. and Mendling, J. (2011). A study into the factors that influence the understandability of business process models. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(3):449–462.

Rodrigues, R. (2013). *Um Framework Genérico para Geração de Texto em Linguagem Natural a partir de Modelos de Processo de Negócio*. Bachelor thesis in Information Systems – Federal University of the State of Rio de Janeiro (UNIRIO).

MAROQ: Um Modelo de Alocação de Recursos Orientado a Qualidade de Experiência

André L. DAmato, Mário Dantas

[1]Departamento de Informática e Estatística – Universidade Federal de Santa Catarina (UFSC)
Santa Catarina – SC – Brasil

{altdamato@gmail.com, mario.dantas@ufsc.br}

Abstract. *The growing trend of cloud computing in network has generated challenges to allocate resources. The quality of experience arises as a important paradigm to solving many challenges. The quality experience takes into account a series of context parameters. Therefore, it is proposed in this paper the resources allocation model MAROQ, which is a quality of experience driven-model that uses context information to resource allocation in clounds, and computational grids. Experimental results show that using context information improves the performance of the task submission.*

Resumo. *A quantidade de recursos fornecidos pelas nuvens computacionais na Internet gerou desafios complexos para resolver problemas relacionados a alocação desses recursos. A qualidade de experiência surge como um paradigma diferenciado como uma fator potencialmente importante na solução desse desafios. A qualidade de experiência leva em consideração parâmetros de contexto. Sendo assim, é proposto neste trabalho o modelo de alocação de recursos MAROQ, que é um modelo orientado a qualidade de experiência, que utiliza informações de contexto para alocação de recursos em nuvens e grids computacionais. Resultados experimentais mostram que utilizar informações de contexto melhora o desempenho na submissão de tarefas.*

1. Introdução

Com a crescente quantidade de nuvens computacionais disponibilizadas na Internet, a utilização de *big data* torna-se um desafio complexo. Esse desafio esta relacionado com a maneira a qual os recursos serão acessados nas nuvens. A modelagem desse tipo de problema muitas vezes esta próxima do uso da teoria do caos [Alkhatib and Krunz 2000] para algumas soluções. Este cenário dificulta a utilização correta dos recursos pelos usuários. A qualidade de um serviço é percebida pelos usuários por fatores cognitivos. Esta percepção subjetiva de qualidade é conhecida como qualidade de experiência (ou em inglês *quality of experience*) (QoE) [Shaikh et al. 2010]. Sendo assim, a maneira com que os recursos são acessados é uma questão determinante para uma boa qualidade de experiência sobre os serviços disponibilizados pelas nuvens computacionais.

Portanto, a qualidade de experiência surge como um paradigma diferenciado e particularmente importante para a alocação de recursos, uma vez que uma boa qualidade de alocação de recursos deve levar em consideração uma serie de parâmetros de contexto [Názario and Dantas 2012] (como por exemplo, tipo de processador, sistema operacional, quantidade de memória). Essas informações de contexto podem ser fornecidas pelo usuário, pois o mesmo conhece o contexto apropriado para execução de sua aplicação.

Sendo assim, este trabalho apresenta o *MAROQ* (Modelo de Alocação de Recursos Orientado a Qualidade de experiência). O modelo proposto visa a alocação de recursos orientada a QoE, considerando o contexto de aplicação. O modelo proposto tem como foco execução de aplicações em nuvens computacionais formadas por *clusters* e *grids*.

O restante do texto é organizado da seguinte maneira: a Seção 2 aborda os principais modelos de QoE propostos nos últimos anos; as Seções 3, e 4 abordam os conceitos básicos para o modelo MAROQ proposto na Seção 5; a seção 6 aborda resultados prévios, e por fim na Seção 7 são apresentadas as considerações finais, e são sugeridas algumas diretivas para trabalhos futuros.

2. Trabalhos Relacionados

Como mencionado anteriormente na seção introdutória, os escalonadores estáticos alocam uma fatia fixa de tempo para um determinado *job*. Diversos modelos como os propostos em [Chase et al. 2003, Sotomayor et al. 2008] utilizam escalonadores estáticos baseados em reserva de recursos. Chase propõe em seu trabalho uma arquitetura flexível para gerenciamento de *grids*. A arquitetura de Chase é considerada flexível por permitir que novos nós sejam configurados dinamicamente de acordo com a disponibilidade dos recursos fornecidos por uma *grid*. No modelo proposto por Chase os recursos são atendidos conforme a demanda. Para garantir que os recursos sejam liberados dinamicamente são realizados monitoramentos nas cargas de trabalho impostas aos conjuntos de nós agrupados logicamente em clusters virtuais (Vclusters) [Chase et al. 2003].

O trabalho de Micillo [Micillo et al. 2009] propõe uma implementação de mecanismos para otimização na alocação de recursos também no contexto de *grids* computacionais. A arquitetura proposta por Micillo possui um gerenciador centralizado, ao qual, é responsável por gerenciar os recursos intergrid. Embora a arquitetura proposta por Micillo suporte migração de jobs, esta funcionalidade depende de compilação utilizando bibliotecas específicas. Hermenier [Hermenier et al. 2010] avança um pouco mais e propõe alocação dinâmica de recursos utilizando uma abstração baseada em transição de estados (executando, esperando, suspenso) para os *jobs*. Assim como Chase e Sotomayor, Hermenier também utiliza máquinas virtuais para compor seu modelo. Neste trabalho foi definida uma abstração de *jobs*, sendo cada *job* entendido como a composição de diversas máquinas virtuais. Estes *jobs* são chamados de *jobs* virtuais. Os *jobs* virtuais possuem 4 estados de execução sendo eles: executando, dormindo, esperando, e terminado. Embora apresentem estratégias para implementação de políticas de alocação de recursos, os trabalhos mencionados anteriormente não proveem mecanismos de contexto para tomada de decisões no momento de alocar recursos. Visando justiça para alocação de recursos, [Ghodsi et al. 2011] propôs em seu trabalho o modelo DFS (*Dominant Resource Fairness*, ou Recursos Dominantes e Justiça) de escalonamento de tarefas para ambientes *MapReduce*. O foco do trabalho de Ghodsi é favorecer o compartilhamento de recursos considerando quais são os recursos dominantes requisitados por uma aplicação. Nesse contexto, recursos dominantes são os recursos que uma determinada aplicação mais depende para executar, por exemplo, aplicações orientadas a processamento demandam mais de CPUs, e memória são prioridade para aplicações orientadas a funções de entrada e saída. No entanto, os recursos não são ajustados com foco nas aplicações, e sim com foco no compartilhamento com o objetivo de obter a máxima utilização dos recursos.

Joseph [Joseph et al. 2012] e Zhou [Zhou et al. 2013] apresentam uma formalização matemática para relacionar alocação de recursos e justiça com qualidade de experiência. O cenário proposto por Zhou remete a redes sem fio multimídia cegas, ou seja, redes sem fio ao qual as estações bases não tem conhecimento sobre a modelagem de qualidade de experiência para o conjunto de aplicações que utilizam o sistema. No entanto ao contrário de Joseph, Zhou não estabelece em sua formalização a relação entre a degradação da qualidade de experiência e a variabilidade da alocação dos recursos com o tempo. Segundo Zhou, a qualidade de experiência varia de acordo com a quantidade de usuários utilizando o sistema e o fator α, que define o grau de justiça para cada usuário quando o mesmo deseja alocar um determinado recurso.

3. Qualidade de experiência e contexto

Existem diferenças de percepção de desempenho sobre um determinado serviço de internet considerando o ponto de vista do usuário e o ponto de vista técnico abordado pelos provedores de serviço [Shaikh et al. 2010]. Isto acontece pelo fato dos provedores utilizarem métrica referentes a qualidade de serviço, ou em inglês *Quality of Service* (QoS), sendo estas métricas estabelecidas a partir de dados técnicos como por exemplo rendimento da rede, vazão, perda de pacotes, atraso etc. Estes parâmetros são geralmente medidos por meio de computadores inseridos na rede que monitoram o desempenho da mesma. No entanto, os usuários percebem o desempenho da rede seguindo métricas mais subjetivas e não técnicas. Esta percepção subjetiva é conhecida como qualidade de experiência, ou em inglês *Quality of Experience* (QoE) [Shaikh et al. 2010].

Para avaliar a percepção do usuário sobre um determinado serviço disponibilizado pela rede, são utilizados questionários aplicados em ambientes controlados. Este tipo de teste é conhecido pela comunidade como pontuação média de opinião, ou em inglês *Mean Opinion Score* (MOS). Esta técnica é geralmente aplicada para avaliar sistemas multimídias. No entanto, devido ao grande número e a diversidade de aplicações testes de opinião não são mais a melhor alternativa. Além disso, os testes de opinião são muito criticados por autores de várias áreas [De Moor et al. 2010]. Estas críticas estão relacionadas a escala de pontuação adotada nos testes, sendo estas consideradas por alguns autores como por exemplo [Koning et al. 2007] como escalas imprecisas e não representativas considerando diferenças culturais de interpretação. Segundo [Sullivan et al. 2008], esta escala determina valores absolutos obtidos em ambientes controlados, ao qual não representam com precisão ambientes reais pelo fato de não considerar a influência de variáveis de contexto.

Contexto é um conjunto de informações que descrevem a situação de uma determinada entidade, que por sua vez pode ser definido como qualquer componente relevante para interação entre uma aplicação e seus respectivos usuário [Dey 2000]. Um sistema é sensível ao contexto (ou em inglês *context-aware*), se este utiliza informações de contexto para prover informações, ou serviços relevantes para o usuário, sendo que a relevância depende diretamente da tarefa exercida pelo usuário dentro do sistema [Dey 2000]. Em outras palavras, as informações de contexto são utilizadas para prover informações, e serviços ajustados as necessidades do usuário. Desta maneira é possível aumentar a satisfação do usuário na utilização do sistema, melhorando consequentemente a qualidade de experiência. Uma boa modelagem de contexto permite também diminuir a complexidade de aplicações computacionais sensíveis ao contexto (context-aware appli-

cations [Schilit et al. 1994]). O trabalho de Bettini [Bettini et al. 2010] é direcionado a computação pervasiva e aborda diversas modelagens de contexto.

4. Alocação de recursos em nuvens

Em ambientes distribuídos os recursos computacionais são determinados por uma cadeia de recursos aleatórios (*network queue theory*). Segundo [Stolyar 2005] os recursos em uma rede computacional necessitam de uma modelagem randômica para utilização dos recursos. Ou seja, os mecanismos de controle de recursos formam uma cadeia de Markov, ao qual as transições de estados não dependem do estado anterior. Isto se deve ao fato da utilização de recursos pelos usuários acontecer de maneira aleatória dentro de uma rede computacional. O escalonamento de recursos é realizado por meio de escalonadores que podem utilizar diferentes políticas de escalonamento. As estratégias mais comuns definidas para os escalonadores de recursos são as seguintes: O FCFS (*First Come First Served*), onde as primeiras tarefas que chegam são as primeiras servidas; *Backfiling* ao qual tenta resolver problemas de fragmentação; Preempção, que melhora os resultados do *backfiling* utilizando informações da aplicação. O método FCFS resulta na fragmentação de recursos, ou seja, diversos recursos se tornam obsoletos no momento da execução das tarefas. O método*backfiling* não gerencia os recursos quando as aplicações mudam seus requisitos dinamicamente. Os mecanismos preemptivos garantem a execução das tarefas de acordo com suas características (por exemplo: prioridade, requisitos, tempo de execução). Para [Joseph et al. 2012], a qualidade de experiência é definida como uma métrica diretamente associada a qualidade de alocação, a variabilidade e a justiça no momento da alocação de recursos. Portanto é possível melhorar a qualidade de experiência de um usuário ajustando estes parâmetros definidos para alocação de recursos.

5. O Modelo MAROQ

O modelo proposto tem como objetivo otimizar a utilização de recursos em nuvens computacionais, visando a satisfação do usuário quanto a utilização dos recursos fornecidos pelo ambiente. O modelo MAROQ visa obter a melhor qualidade de experiência resultante a partir do gerenciamento dos parâmetros: significância, variabilidade e justiça na alocação de recursos. O plano formado pelos parâmetros citados determinam a qualidade de experiência [Joseph et al. 2012]. Neste sentido, é proposto que o modelo MAROQ seja orientado a contexto para realizar a melhor alocação de recursos possível ajustando assim o parâmetro de significância para que os recursos não precisem ser realocados durante a execução de uma aplicação. Além disso, o contexto permite a otimização do espaço de busca por recursos melhorando assim o desempenho para realizar a alocação dos mesmos.

Como mencionado anteriormente, a QoE em um sistema computacional segundo [Joseph et al. 2012] é a composição dos parâmetros alocação ajustada *mean*, variabilidade e justiça. Seguindo esta definição é proposto no modelo MAROQ a utilização de contexto para lidar com a alocação ajustada e com a variabilidade na alocação de recursos. Em outras palavras, a utilização de contexto permite alocar recursos de maneira que os mesmos não precisem ser realocados posteriormente. A justiça na alocação de recursos é realizada a partir de preempção de tarefas. O mecanismo de preempção garante que uma tarefa de menor prioridade, libere os seus recursos no momento que uma tarefa de maior prioridade chega.

O *middleware* [Scheidt et al. 2012] implementado no modelo é composto por um mecanismo que aloca recursos de acordo com a prioridade de cada tarefa, lidando assim com o parâmetro justiça. A Figura 1 ilustra o modelo MAROQ. O plano de controle resultante visa atuar como gerenciador de clusters em nuvens computacionais, focando sempre em obter a melhor QoE possível.

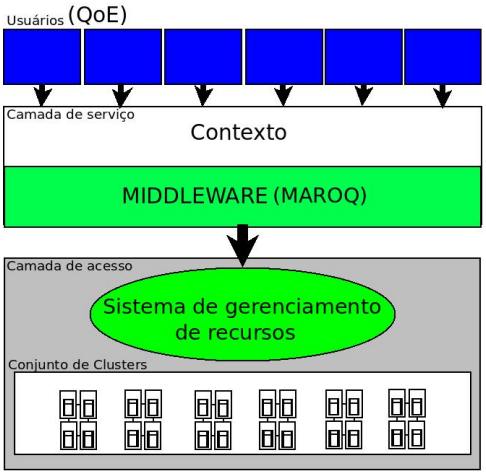


Figura 1. Modelo proposto

Na Figura 1 a parte superior ilustra os usuários dos recursos disponíveis pelos clusters, sendo estes responsáveis pela requisição dos recursos inserindo informações de contexto. As informações de contexto são referentes ao tipo de processador, sistema operacional, quantidade de memória necessária. A partir destas informações o MAROQ imediatamente realiza a síntese de um arquivo de submissão, que é utilizado como entrada para o sistema de gerenciamento de recursos. O modelo utiliza o balanceador de carga baseado no próprio gerenciador de recursos.

Uma outra informação de contexto importante considerada no modelo é inerente a quantidade de comunicação entre os processos, pois esta informação revela a necessidade de desempenho dos componentes de interconexão intra e inter-cluster. Ou seja, quanto maior a quantidade de comunicação entre os processos da aplicação, maior deverá ser a velocidade do componente(s) de interconexão do(s) cluster(s) alocado(s). Pois a latência de comunicação é menor entre processos dentro de um mesmo cluster, do que em processos alocados em clusters geograficamente distribuídos. Por outro lado, se uma aplicação é composta por processos que não se comunicam esta pode ser alocada em clusters distintos. A Figura 2 mostra a arquitetura do sistema MAROQ. A parte superior da figura representa o usuário, que realiza a requisição de execução de uma aplicação passando o contexto desejado para esta execução. Uma vez determinada o contexto e realizada a requisição esta é enviada via *web service* [Kalin 2013] utilizando o protocolo SOAP [Ryman 2001]. A requisição realizada pelo usuário é recebida pelo meta escalonador do MAROQ. O meta escalonado realiza assim uma negociação com os lideres de cada *cluster* a fim de identificar o que possui o melhor conjunto de recursos para atender ao contexto especificado. Após o *cluster* que atenda aos requisitos ser identificado, o líder deste cluster responsável por executar o controle do gerenciamento de recursos recebe via *web service* os dados para execução da tarefa a ser submetida. Os nós de cada *cluster* são responsáveis apenas por executar as aplicações.

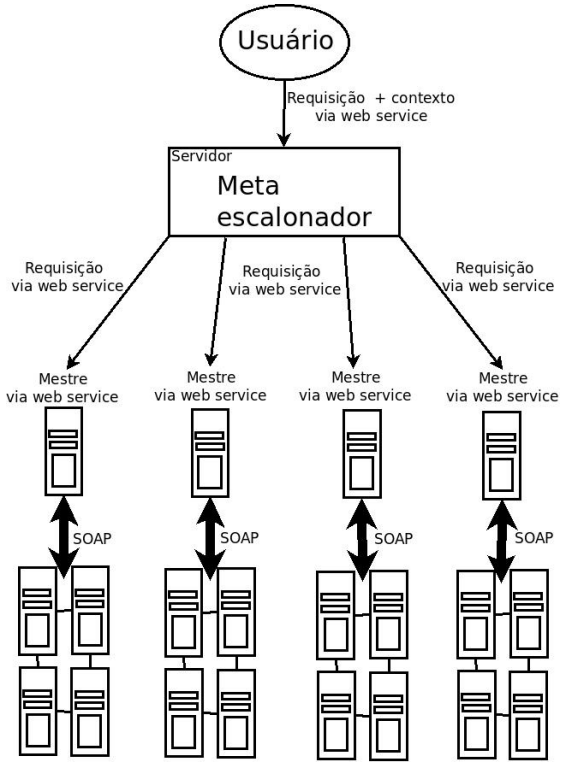


Figura 2. Arquitetura do MAROQ

6. Resultados

O gráfico da Figura 3, mostra que as submissões que utilizam informações de contexto possuem maior desempenho em relação a tempo, quando comparadas as submissões que não utilizam contexto.

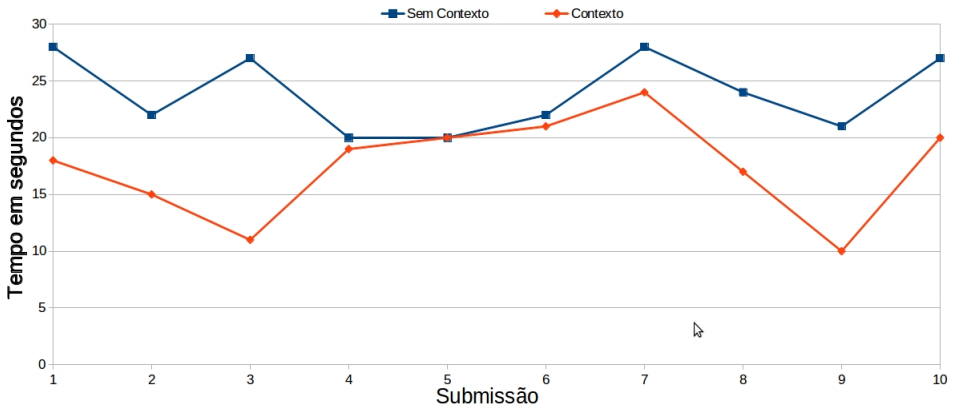


Figura 3. Resultados

O experimento foi realizado em 10 máquinas virtuais com 1 núcleo cada uma utilizando a ferramenta VmWare. O sistema de gerenciamento utilizado neste experimento foi o Condor [Frey et al. 2001]. A aplicação submetida é composta por um *sleeper*, que recebe o tempo que a aplicação ficará dormindo como parâmetro na invocação do programa. Para todas as submissões o parâmetro utilizado foi o mesmo. As informações

de contexto consideradas foram as seguintes: número de cpus, quantidade de memória, tamanho de disco rígido e sistema operacional. A aplicação foi escolhida pelo fato da utilização da função *sleep* permitir isolar os tempos gastos com a aplicação e com a alocação de recurso, sendo que é conhecido o tempo despendido com a função *sleep*. No caso da aplicação submetida ao experimento o parâmetro utilizado foi 5 segundos. Ou seja, dos tempos mencionados no gráfico da Figura 3 é possível dizer que aproximadamente 5 segundos foram gastos com a função *sleep*, sendo o tempo restante utilizado para as demais rotinas da aplicação e do sistema de alocação de recursos. Os resultados obtidos com o experimento mostram que inserir informações de contexto diminui o tempo para realizar a submissão de tarefas. Isto ocorre pela diminuição do espaço de busca por recursos.

7. Conclusão e Trabalhos Futuros

Este trabalho apresenta o modelo MAROQ, sendo este um modelo de alocação de recursos para nuvens computacionais orientado a QoE e contexto. Informações de contexto permitem que os recursos sejam ajustados as necessidades da aplicação. Desta maneira é possível melhorar o desempenho do sistema e consequentemente melhorar a QoE do usuário. Os experimentos realizados revelaram que o modelo permite melhorar o desempenho na alocação de recursos. Isso ocorre pelo ajuste e diminuição do espaço de busca de recursos. Como as informações de contexto ajustam os recursos de acordo com a aplicação, acredita-se que o MAROQ possua potencial para melhorar também o desempenho quanto ao tempo de execução. Sendo assim, é pretendido como trabalho futuro realizar experimentos focando no tempo de execução utilizando informações de contexto.

Referências

Alkhatib, A. and Krunz, M. (2000). Application of chaos theory to the modeling of compressed video. In *Communications, 2000. ICC 2000. 2000 IEEE International Conference on*, volume 2, pages 836–840 vol.2.

Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive Mob. Comput.*, 6(2):161–180.

Chase, J. S., Irwin, D. E., Grit, L. E., Moore, J. D., and Sprenkle, S. E. (2003). Dynamic virtual clusters in a grid site manager. In *Proceedings of the 12th IEEE International Symposium on High Performance Distributed Computing*, HPDC '03, pages 90–, Washington, DC, USA. IEEE Computer Society.

De Moor, K., Ketyko, I., Joseph, W., Deryckere, T., De Marez, L., Martens, L., and Verleye, G. (2010). Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting. *Mobile Networks and Applications*, 15(3):378–391.

Dey, A. K. (2000). *Providing Architectural Support for Building Context-aware Applications*. PhD thesis, Atlanta, GA, USA. AAI9994400.

Frey, J., Tannenbaum, T., Livny, M., Foster, I., and Tuecke, S. (2001). Condor-g: a computation management agent for multi-institutional grids. In *High Performance Distributed Computing, 2001. Proceedings. 10th IEEE International Symposium on*, pages 55–63.

Ghodsi, A., Zaharia, M., Hindman, B., Konwinski, A., Shenker, S., and Stoica, I. (2011). Dominant resource fairness: Fair allocation of multiple resource types. In *NSDI*, volume 11, pages 24–24.

Hermenier, F., Lèbre, A., and Menaud, J.-M. (2010). Cluster-wide context switch of virtualized jobs. In *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, HPDC '10, pages 658–666, New York, NY, USA. ACM.

Joseph, V., de Veciana, G., and Arapostathis, A. (2012). Resource allocation: Realizing mean-variability-fairness tradeoffs. In *Allerton Conference*, pages 831–838. IEEE.

Kalin, M. (2013). *Java Web Services: Up and Running*. O'Reilly Media, Inc.

Koning, T. C. M., Veldhoven, P., Knoche, H., and Kooij, R. E. (2007). Of MOS and men: bridging the gap between objective and subjective quality measurements in mobile tv. In *SPIE*.

Micillo, R., Venticinque, S., Aversa, R., and Di Martino, B. (2009). A grid service for resource-to-agent allocation. In *Internet and Web Applications and Services, 2009. ICIW '09. Fourth International Conference on*, pages 443–448.

Názario, D. C. and Dantas, M. A. R. (2012). Taxonomia das publicações sobre qualidade de contexto. In *International Journal of Sustainable business*, number 20, pages 1–28.

Ryman, A. (2001). Simple object access protocol SOAP and web services. In *Proceedings of the 23rd International Conference on Software Engineering*, ICSE '01, pages 689–, Washington, DC, USA. IEEE Computer Society.

Scheidt, R. F., Schmidt, K., Pessoa, G. M., Viera, M. A., and Dantas, M. A. R. (2012). A software product line approach to enhance a meta-scheduler middleware. In *Journal of Physics: Conference Series*, volume 341, pages 1–7.

Schilit, B., Adams, N., and Want, R. (1994). Context-aware computing applications. In *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications*, WMCSA '94, pages 85–90, Washington, DC, USA. IEEE Computer Society.

Shaikh, J., Fiedler, M., and Collange, D. (2010). Quality of experience from user and network perspectives. *annals of telecommunications*, 65(1-2):47–57.

Sotomayor, B., Keahey, K., and Foster, I. (2008). Combining batch execution and leasing using virtual machines. In *Proceedings of the 17th International Symposium on High Performance Distributed Computing*, HPDC '08, pages 87–96, New York, NY, USA. ACM.

Stolyar, A. L. (2005). Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Syst. Theory Appl.*, 50(4):401–457.

Sullivan, M., Pratt, J., and Kortum, P. (2008). Practical issues in subjective video quality evaluation: Human factors vs. psychophysical image quality evaluation. In *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video*, UXTV '08, pages 1–4, New York, NY, USA. ACM.

Zhou, L., Chen, M., Qian, Y., and Chen, H.-H. (2013). Fairness resource allocation in blind wireless multimedia communications. *Multimedia, IEEE Transactions on*, 15(4):946–956.

Paralelização de Autômatos Celulares em Placas Gráficas com CUDA

Marcos Paulo Riccioni de Melos [1]

[1]Departamento de Ciência da Computação – Instituto Multidisciplinar – Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Nova Iguaçu – RJ – Brasil

marcospaulo.r.melos@gmail.com

Abstract. *Cellular automata is a technique for solving problems containing elements that interact with their immediate vicinity and requiring external data instance. The efficiency of the algorithm was tested in several scenarios and compared between the sequential and parallel approaches. The results show that the methodology used to improve the performance problems of this nature, may be improved by using the shared memory of the GPU. There is also a limitation on the size of the problem in the parallel implementation with this distribution in more plates and more computers would be a more accurate solution.*

Resumo. *Autômatos celulares é uma técnica de solução de problemas contendo elementos que interagem com sua vizinhança próxima e necessitam de dados externos a instância. A eficiência do algoritmo foi testada em diversos cenários e comparadas entre as abordagens sequencial e paralela. Os resultados mostram que a metodologia usada melhora o desempenho para problemas dessa natureza, podendo ser melhorada com o uso da memória compartilhada da GPU. Também há uma limitação no tamanho do problema na implementação paralela, com isso a distribuição em mais placas e mais computadores seria uma solução mais acertada.*

1. Introdução

A abordagem de autômatos celulares consiste em uma grelha infinita e regular de células, cada uma podendo estar em um número finito de estados, que variam de acordo com regras determinantes. Esta técnica se adapta muito bem a problemas que discretizam populações e simulam a interação extracelular das mesmas, criando vínculos de vizinhança que são atualizados a cada interação de tempo na população [SCHIFF 2008].

O tempo também é discretizado, onde o estado de uma célula no tempo t é obtido em função do estado no tempo t-1, levando em consideração seu estado e os estados de células vizinhas. Esta vizinhança corresponde a células próximas que obedecem a uma formação pré-definida. Geralmente, os problemas são discretizados em forma de matrizes, no qual cada elemento é uma ou mais posições da matriz [SCHIFF 2008; GREMONINI & VICENTINI 2008].

[GARDNER 1970] e [SCHIFF 2008] apresentam a aplicação teórica Gol (Game of Life). A mesma reproduz, através de regras simples, as alterações e mudanças em

grupos de seres vivos, ilustrando o comportamento de células em relação ao tempo e seu ambiente.

O Gol apresenta apenas quatro regras definidas por [Gardner 1970].

O custo para calcular os estados das células e obter as alterações e mudanças de grupo pode ficar alto, dependendo do tamanho da matriz e da quantidade de gerações a serem simuladas. O objetivo deste trabalho é melhorar o desempenho através da paralelização massiva dos passos da aplicação. Para tanto, foi desenvolvida uma versão paralela do GoL, denominada GolCU, para aproveitar o paralelismo disponível nas placas gráficas (GPU – Graphics Processor Unit) através da linguagem CUDA (Compute Unified Device Architecture), que devido a sua natureza SIMD (Single Instruction Multiple Data) mostra-se muito adequada a problemas desta classe, onde obtém um ganho significativo em performance frente a processadores convencionais, pois fornece quantidades superiores de unidades de processamento [EICHENBERGER et al., ____; SANDERS & KANDROT 2010].

2. Metodologia

A solução paralela abordada foi utilizando apenas uma GPU. Outras abordagens também podem ser criadas, tendo em vista a natureza do problema.

A aplicação GoL é inerentemente paralela e muito adequada para executar em GPU, já que é explícito no tempo. A principal característica da GPU é processar a mesma instrução sobre diferentes dados, logo, cada core da GPU pode executar uma célula da matriz utilizada na simulação do GoL. Assim, na abordagem proposta, os dados são copiados para a memória global da GPU e cada thread calcula paralelamente o estado de sua célula correspondente em um instante de tempo. O resultado final da simulação é copiado de volta para CPU para exibição da simulação.

A quantidade de threads criadas depende do tamanho do problema e da quantidade de cores disponíveis na GPU. Quando o tamanho da aplicação, ou seja, quando a quantidade de threads ou células é maior do que a quantidade de cores disponíveis, o processamento continua sendo feito em paralelo, mas é criada uma fila para o processamento das células excedentes. Além disso, outra característica da aplicação é que a cada passo, é necessário a sincronização das células para que todas as threads reiniciem o processamento com o dado atualizado.

No caso ideal para a implementação apresentada e o ambiente de execução disponível, o tamanho máximo da matriz de entrada seria de aproximadamente 8191x8191, ou um total de 67.107.840 elementos, de acordo com a capacidade máxima da placa utilizada. Essa capacidade máxima é definida pela quantidade de blocos e quantidade máxima de threads por bloco disponíveis na GPU.

3. Implementação

Utilizando o Visual Studio 2010, a implementação foi baseada na linguagem C/C++, utilizando o CUDA como arquitetura de computação paralela.

A parte lógica do algoritmo Gol é mantida para todas as abordagens sugeridas. Por ser uma aplicação teórica, não há entrada nem saída esperada correta, apenas a correta computação até a última interação do tempo.

3.1. Gol Sequencial

A seguir é exemplificada a parte principal do código do *Gol* sequencial:

```
repita para i menor que geracao {
        repita para j menor que tamX
            repita para k menor que tamY {
                    regras do Gol;
            }
}
```

Como visto, essa abordagem requer laços de repetição aninhados, o que ocasiona grande gasto de tempo apenas para as atualizações das matrizes.

3.2. Gol paralelo com 1 GPU

A seguir são exemplificados alguns trechos do código do Gol paralelo:

```
__global__ void calcViz(const int h_tam, int *d_matrix, int *d_aux){
        regras do Gol;
        __syncthreads();
}
int main(){
        repita i menor que geracao {
            calcViz <<<dimGrid, dimBlock>>> (h_tam, d_matrix, d_aux);
            cudaStatus = cudaDeviceSynchronize();
        }
}
```

Como dito anteriormente, nesta abordagem todos os dados são transferidos para a GPU, que calcula cada geração de modo paralelo dentro da placa. Cada thread é responsável pelo cálculo do estado de uma célula.

É passado para a função do Device o tamanho do vetor, o vetor com os dados iniciais da simulação, além de um vetor auxiliar. Cada célula é identificada pelo identificador único da thread, nessa divisão cada thread realiza os cálculos referente a célula correspondente do vetor de entrada. Após, os dados são atualizados no vetor principal, para serem usados na próxima interação.

5. Resultados e Discussão

Para avaliar a proposta foi utilizado o seguinte ambiente: processador Intel Core i7 3770, com clock de 3,4GHz, 8mb de memória cache e 12Gb de memória RAM a 1333MHz, 2 placas de vídeo NVIDIA GT 640, Ubuntu 12.04, CUDA 6.

Os resultados são apresentados na Tabela 1, onde foram utilizados valores médios a partir de cinco tomadas de tempo. Os tempos das implementações foram obtidos utilizando a função clock() da linguagem C, onde é medido o tempo de toda a implementação e na implementação paralela foi usada cudaEventRecord().

O sistema operacional foi colocado em modo texto e nenhum outro processo foi inicializado.

Table 1. Tempos medidos

Tamanho da Matriz	Quantidade de de Gerações	Tempo médio sequencial (ms)	Desvio padrão tempo sequencial(ms)	Tempo médio paralelo (ms)	Desvio padrão tempo paralelo(ms)	Speed-Up
10	1000	10,000	0,000	21,207	0,776	0,4715
100	1000	480,000	0,000	170,221	0,801	2,8198
1000	1000	48.260.000,000	5,477	13.474.863,281	18,114	3,5814
2000	1000	192.430.000,000	11,401	53.914.535,156	50,182	3,5691
5000	1000	1.199.970.000,000	91,268	336.979.875,000	64,591	3,5609
8000	1000	3.072.020.000,000	179,220	864.370.187,500	65,320	3,5540
10000	1000	4.800.080.000,000	208,989	1.352.506.500,000	76,621	3,5490

O speed-up foi utilizado como medida de comparação entre os tempos de execução sequencial e paralelo [HENNESSY & PATTERSON 2011]. Foi observado que o tempo paralelo em relação a matriz quadrada de tamanho 10 é pior que o sequencial, isso é explicado pela necessidade de troca de informações entre o Host e Device. Quando aumenta-se o tamanho da matriz a perda inicial com a comunicação é absorvida pelo tempo da computação, com isso o desempenho do sequencial é facilmente ultrapassado.

6. Conclusão

A partir dos resultados foi possível observar que, problemas resolvidos por autômatos celulares se beneficiam com a abordagem apresentada. A implementação paralela desenvolvida pode ser melhorada em com a utilização da memória compartilhada. A limitação de tamanho pode ser aumentada repartindo os dados para cálculos em diversos computadores, com diversas GPU's ou atribuindo a computação de mais células a cada thread.

7. Referências

Gardner M. (1970). The fantastic combinations of John Conway's new solitaire game "life". Scientific American, v. 223, n. 4, p. 120-123.

Schiff J. L. (2008) Cellular Automata: A discrete View of the World. Wiley Interscience, New Jersey.

Sanders J. and Kandrot E. (2010) CUDA by Example: an introduction to general-purpose GPU programming. Addison-Wesley Professional.

Eichenberger A. E., Wu P. and O'Brien K. Vectorization for SIMD Architectures with Alignment Constraints. IBM T.J. Watson Research Center, Yorktown Heights, NY. Disponível em < http://researcher.watson.ibm.com/researcher/files/us-alexe/paper-eichen-pldi04.pdf >. Acesso em julho de 2014.

Gremonini L. and Vicentini E. (2008) Autômatos celulares: revisão bibliográfica e exemplos de implementações. Revista Eletrônica Lato Sensu – UNICENTRO, ed. 6.

Hennessy J. L. and Patterson D. A. (2011) Computer Architecture: A quantitative Approach. Morgan Kaufman, 5th ed.

Detecção de Requeima em Tomateiros Apoiada Por Técnicas de Agricultura de Precisão

Diogo Nunes[1], Carlos Werly[1], Pedro Vieira Cruz[1], Marden Manuel Marques[3] Sérgio Manuel Serra da Cruz [1,2,4]

[1]Universidade Federal Rural do Rio de Janeiro/UFRRJ
[2] Programa Pós Graduação em Modelagem Matemática e Computacional/UFRRJ
[3] Centrais de Abastecimento do Estado do Rio de Janeiro – CEASA/RJ
[4] Programa PET Sistemas de Informação - PET-SI/UFRRJ

{diogo_c_nunes, carlos_werly, serra}@ufrrj.br

Resumo. *Cada vez mais os produtores percebem que a tomada de decisões e uso de sistemas inteligentes na agricultura é mais que uma tendência, torna-se uma questão de sobrevivência e necessidade devido a globalização da economia e a competitividade dos produtos agrícolas. Este trabalho apresenta uma abordagem focada na melhoria da qualidade de culturas de tomate. Desenvolvemos estratégias computacionais de baixo custo voltadas para apoiar agricultores familiares na detecção precoce da requeima. Nossa abordagem utiliza técnicas do domínio da Agricultura de Precisão, sendo aplicadas em cultivares de tomates de um campo experimental onde foram processadas observações de campo, imagens e anotações coletadas pelos agricultores e utilizadas em redes neurais para a detecção da doença.*

1. Introdução

O Brasil possui uma agricultura familiar dinâmica e diversificada, composta por 4,3 milhões de pequenos estabelecimentos agrícolas, responsáveis pela produção de produtos importantes para a cesta básica. Esta diversidade representa um valor econômico enorme para a agricultura brasileira que hoje experimenta um forte ritmo de crescimento em sua produtividade. Em 2013, a agricultura familiar contribuiu com cerca de 6% do Produto Nacional Bruto do Brasil [IBGE 2013]. Na área agrícola, as tecnologias da informação e comunicação (TICs) são os recursos e suportes tecnológicos que permitem o fluxo de informações e abrangem diversos meios de comunicação, desde os mais antigos como rádio, televisão, telefones fixos até mais modernos, como telefones celulares, computadores, *tablets*, equipamentos de gravação de áudio e vídeo, redes e sistemas multimídia, entre outros. As TICs conferem algum tipo de empoderamento às comunidades de pequenos agricultores [Schwartz 2012]. De acordo com o Comitê Gestor da Internet [CGI 2011], a partir do ano de 2010, os lares da zona rural apresentaram o maior crescimento de posse de telefone celular, passando de 58 % em 2009 para 68 % em 2010. No entanto, a agricultura familiar no estado do Rio de Janeiro pouco se utiliza de aplicativos móveis para telefones celulares ou *tablets* para auxiliá-los na gestão de suas lavouras [Nunes et al. 2014].

Dentre as culturas temporárias com valor expressivo na cadeia produtiva do estado do Rio de Janeiro destacam-se o tomate (*Solanum lycopersicum*). O tomate é a mais importante dentre todas as hortaliças cultivadas no Brasil e o Rio de Janeiro é hoje o terceiro maior produtor nacional. Os tomates são frutos climatérios e altamente suscetíveis a doenças e contaminação [Nakano 1999]. Além disso, o uso indiscriminado de agrotóxicos nos tomateiros pode acarretar sérios problemas à saúde humana e ao meio ambiente [Correa et al. 2009]. Por último, mas não menos importante, os pequenos agricultores familiares podem não ter os recursos necessários para cumprir os padrões cada vez mais rigorosos de segurança

alimentar, como a rastreabilidade, verificação de segurança e controle de estoque, principalmente porque eles não possuem pleno acesso a aplicativos móveis que podem auxiliar no alerta ou detecção precoce da ocorrência de doenças nos tomateiros [Vianna e Cruz 2013; Nunes et al. 2014]. Essa classe de aplicativos móveis pode ajudar os pequenos produtores e agricultores a reduzir o manejo de produtos químicos nas doenças dos tomateiros, aumentar seus níveis de renda e oferecer produtos mais saudáveis.

Este trabalho tem como objetivo apresentar um ambiente computacional que auxilia na detecção precoce de doenças foliares em tomateiros. O ambiente se alinha com a temática da Agricultura de Precisão (AP) [Coelho e Silva 2011; MAPA 2014], sendo capaz de manipular dados sobre a cultura auxiliando pequenos agricultores a acompanhar o desenvolvimento das culturas e no planejamento de atividades mais sustentáveis. Dentre as tecnologias computacionais adotadas temos inteligência computacional sob a forma de reconhecimento de padrões baseados em redes neuronais tipo *Multilayer Perceptron* (MLP) [Sanyal e Patel 2008] e a engenharia do conhecimento aplicada em sistemas inteligentes em agricultura [Quincozes et al 2010] com ênfase na gestão de dados e seus descritores. O uso integrado dessas tecnologias já está se tornando uma realidade em grandes propriedades [Wang et al. 2006], contudo seu uso em pequenas propriedades e lavouras de tomates é uma novidade bastante desafiadora, ainda mais se considerarmos as dificuldades envolvidas na detecção precoce e automatizada de doenças e as dificuldades de aceitação da nova prática pelos pequenos produtores.

Este trabalho está organizado da seguinte forma, a Seção 2 caracteriza a hortaliça e a requeima. A seção 3 apresenta os conceitos da AP utilizados neste trabalho. A seção 4 apresenta a arquitetura para coletar dados da cultura e efetuar o reconhecimento dos padrões da doença. A seção 5 apresenta detalhes do banco de dado e experimentos baseados em redes neurais. A seção 6 apresenta os trabalhos relacionados, finalmente, a seção 7 conclui o artigo.

2. Requeima em Tomateiros

Doenças de plantas são anomalias causadas por fatores bióticos ou abióticos que agem continuamente na planta, resultando em queda de produção e/ou perda da qualidade do produto [DISQUAL 2010]. No tomateiro, as doenças são de frequência ou intensidade variadas em função de fatores ou condições (clima, localização da área plantada, modo de implantação e de condução da lavoura) [Filgueira 2008]. Já foram relatadas mais de 200 doenças em tomateiros, sendo que a prevalente no Brasil é a requeima. Ela é causada pelo oomycetes denominado *Phytophthora infestans*, ela é severa e resulta em perdas significativas nas culturas. A doença é visualmente reconhecida pelo surgimento de pontos escuros nas folhas, cujos matizes variam do cinza ao verde-pálido, frequentemente localizadas nas bordas da folha, podendo evoluir para grandes áreas necrosadas marrons [Filgueira 2008]. Essas lesões causam a perda de folhas e, nos casos mais severos, a morte da planta. Apesar de tipicamente observados nas folhas, os sintomas também podem aparecer em caules, frutos e brotos [Correa 2009].

3. Agricultura de Precisão

A Agricultura de Precisão (AP) é um tema multidisciplinar. É um sistema de manejo integrado de informações e tecnologias de *hardware* e *software*, baseado nos conceitos de que as variabilidades de tempo e espaço influenciam nos rendimentos das lavouras. AP visa o gerenciamento planejado do sistema de produção agrícola como um todo, não só das aplicações de insumos ou de mapeamentos. AP faz uso de um grande conjunto de tecnologias,

a saber: GPS, SIG, sistema de informações, bancos de dados, sensores para medidas ou detecção de parâmetros ou de alvos de interesse no agroecossistema (solo, planta, insetos e doenças), de geoestatística e da mecatrônica [MAPA 2014].

AP não se relacionada apenas com uso de TICs, seus fundamentos podem ser empregados no dia-a-dia das propriedades pela maior organização, gestão e controle das atividades, dos gastos e produtividade em cada área cultivada. A diferenciação da produção já ocorre na divisão e localização das lavouras dentro da propriedade rural, na divisão dos talhões, ou simplesmente, na identificação de "manchas" que diferem do padrão geral. A partir dessa divisão, o tratamento diferenciado de cada área é a aplicação das técnicas de AP [Coelho e Silva 2011]. A Figura 1 ilustra os principais conceitos da AP utilizados na modelagem dos aplicativos móveis, Web, redes neurais e do banco de dados.

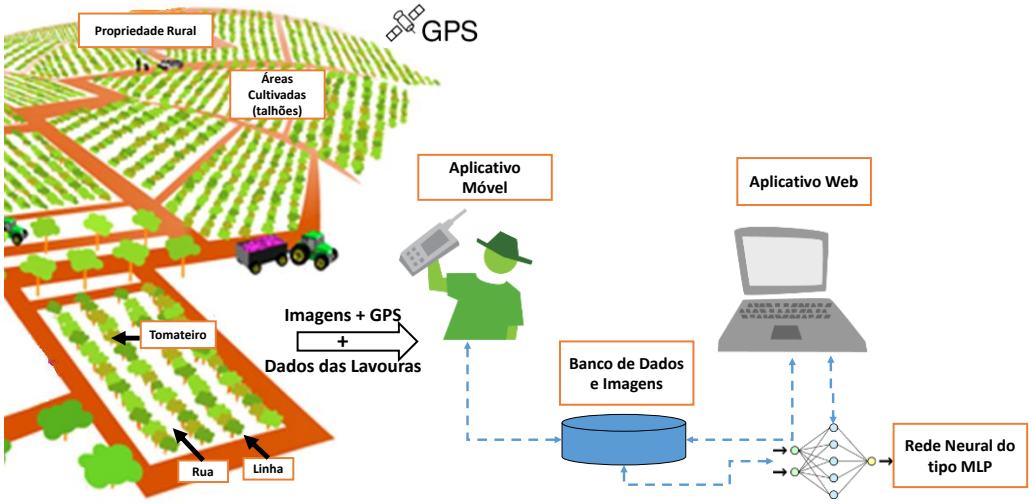


Figura 1 - Visão geral dos conceitos da AP utilizados no manejo tecnológico dos tomateiros.

4. Arquitetura Proposta

A arquitetura proposta está em fase de finalização do desenvolvimento, ela é composta por camadas de aplicativos Móvel e Web cujas funcionalidades foram elicitadas através de técnicas de análise dos requisitos levantados junto aos agricultores e agrônomos e a modelagem foi representada em linguagem UML. O *schema* do banco de dados relacional foi concebido para armazenar os dados sobre as lavouras, áreas cultivadas, plantas e suas imagens, dados de GPS, procedimentos agrícolas e também metadados sobre o monitoramento realizado pelos agricultores através dos dispositivos móveis. O aplicativo móvel (Figura 2) possui interfaces gráficas (bem simples) para serem utilizadas pelos agricultores com pouca experiência em TICs. Através dele o agricultor é capaz de representar as áreas cultivadas (ruas e linhas), coletar dados, anotações e imagens sobre cada tomateiro. Estes dados são transferidos através da conexão sem fio para uma estação base que executa o aplicativo Web. O aplicativo Web é capaz de manter as propriedades rurais, áreas de cultivo, lavouras, mão de obra, aplicações de defensivos, anotações de cultivo e estados meteorológicos. Ele também é capaz de importar os dados da aplicação móvel e produzir os relatórios gerenciais consolidados de acompanhamento a serem submetidos aos serviços de assistência técnica. O aplicativo Web também desempenha um papel importante pois é

através dessa interface que ocorre que executam as tarefas de reconhecimento de padrões para a detecção automatizada de requeima baseadas em redes neuronais.

5. Banco de Dados e Detecção Automatizada de Requeima

O aplicativo móvel foi desenvolvido em Java (Figura 2), sendo baseado no *Android* (versão *KitKat*), opera em simples telefones celulares e não requer que a área cultivada seja atendida pela rede de telefonia celular. O aplicativo é capaz de coletar e armazenar imagens georreferenciadas sobre os tomateiros infectados (ou suspeitos) através uma câmera embutida. Além disso, permite que o agricultor registre e identifique unicamente cada tomateiro da propriedade através do retículo Rua-Linha-Talhão. Cada tomateiro possui uma identificação única, descritores e imagens que representam seu estado geral que podem variar ao longo do tempo desde o plantio até a colheita dos frutos.

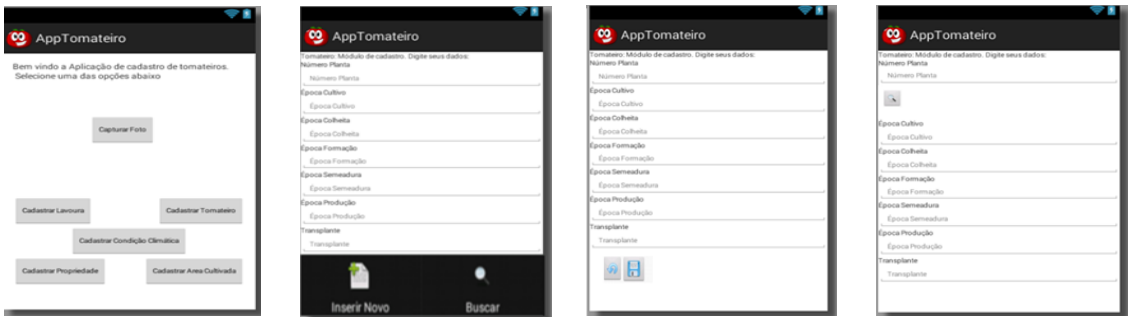


Figura 2. Telas do Aplicativo Móvel no *Android*.

Os dados coletados pelo aplicativo móvel são armazenados temporariamente no SQLite. O SQLite é uma biblioteca que implementa um banco de dados transacional compatível com SQL de modo embutido no dispositivo móvel. Posteriormente os dados são transferidos para o SGBD MySQL. O *schema* do banco de dados foi idealizado para armazenar informações textuais (sobre os conceitos da AP), cultivo, imagens e metadados (Figura 3). O *schema* foi materializado para representar os atributos que são utilizados pele aplicativos e pela rede neural MLP que faz a detecção automatizada da ocorrência de requeima.

5.1 Banco de Dados

A Figura 3 ilustra apenas as principais entidades do *schema* do banco de dados necessárias à compreensão deste trabalho. *Propriedade Rural* é uma localidade na micro-região estudada, contém descritores sobre fatores intrínsecos da mesma, como qualidade da água, topografia predominante, mecanização e seu tipo. Já o relacionamento *Área Cultivada* armazena dados sobre o cultivo realizado, assim como, a estrutura da área e a qualidade do solo pertencente ao local. *Rua* trata do dimensionamento e espaço físico que o produtor rural tem para se locomover entre as linhas e para manipular os tomateiros, cada rua possui duas *Linhas* de tomateiros, sendo que este relacionamento terá a quantidade de linhas e a distância entre as mesmas. *Condição Climática* armazena as características climáticas predominante da propriedade (luminosidade, temperatura, umidade) influenciam significativamente na cultura do tomate. *Lavoura* descreve a técnicas de cultivo e os genótipos (variedades) de tomate, dados da semente, assim como a quantidade de plantas da lavoura, entre outros. *Praga, Doença e Injuria* contém informações referentes ao controle de insetos e ácaros (pragas) e doenças dos tomateiros. Injuria são pequenas lesões ocasionadas na planta que cria condições propícias ao surgimento dos itens citados anteriormente.

A entidade *Tomateiro* armazena os dados específicos do ciclo de vida dos tomateiros (épocas de cultivo, colheita, formação, semeadura, produção e transplante). Todos estes itens citados juntamente com a entidade *Imagem* são utilizados pelas redes neurais. Cada imagem de tomateiro pode ser georeferenciada permitindo detectar automaticamente a requeima e também avaliar o deslocamento da doença pelos talhões e pela propriedade.

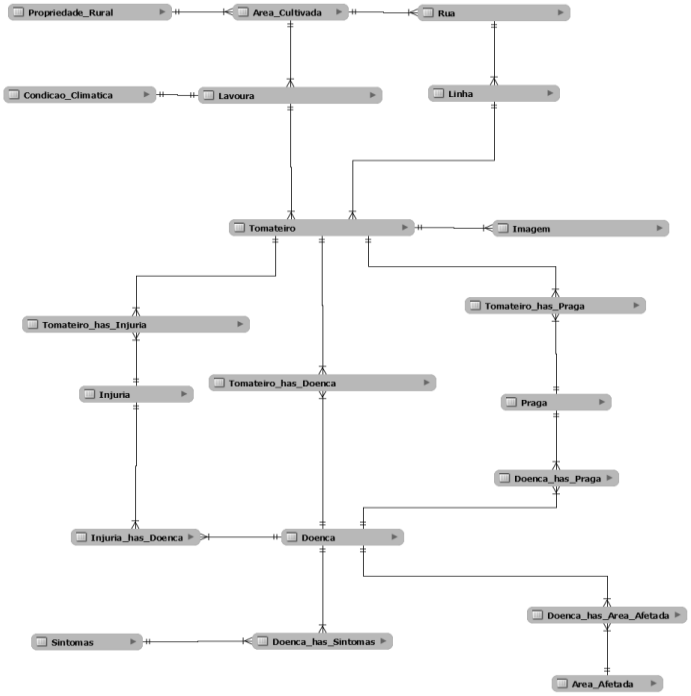


Figura 3. *Schema* resumido do banco de dados da abordagem proposta.

5.2 Detecção da Requeima

Até o momento, os agricultores familiares dispõem de poucas opções de ferramentas automatizadas que os auxiliem na detecção precoce da requeima nos tomateiros. Uma abordagem utilizada é a chave de classificação manual definida por Correa et al (2009). Essa chave está baseada em imagens estilizadas de folhas de tomates que quantificam o grau de infestação por *P. infestans,* permitido avaliar visualmente o grau de contaminação do tomateiro. A detecção manual precoce de doenças, realizada pelos especialistas, nem sempre é uma opção economicamente viável. Por outro lado, a detecção automatizada acelera o processo de identificação das amostras ao se separar em dois conjuntos: aquele das amostras em bom estado e aquele das amostras que podem passar por um exame mais detalhado, a ser feito por especialistas. A incorporação de técnicas inteligência computacional tem como objetivo automatizar e acelerar o processo de identificação de doenças e pragas em tomateiros.

Recentemente, Vianna e Cruz (2013, 2014) desenvolveram técnicas que podem incrementar a produtividade das lavouras de tomates do estado do Rio de Janeiro. A tecnologia computacional adotada foi baseada na inteligência computacional, sob a forma de reconhecimento de padrões baseados em redes neurais do tipo MLP. Nestes trabalhos, os autores realizaram dois processamentos sobre cada amostra digital colorida de tomateiros: a aplicação de um filtro vermelho/verde e a conversão para imagem em tons de cinza. Sobre cada amostra, foram realizadas contagens de pixels que se encontram dentro de faixas de

cores ou tons de cinza relevantes. O conjunto de amostras da investigação perfaz um total de 226 imagens digitais obtidas de amostras de 66 tipos diferentes de genótipos plantados em campos experimentais do setor de horticultura do Departamento de Fitotecnia do IA/UFRRJ, uma área historicamente associada com a ocorrência natural da requeima. Os genótipos incluem um cultivar comercial suscetível à doença, dois resistentes e 63 ainda sob avaliação na Universidade.

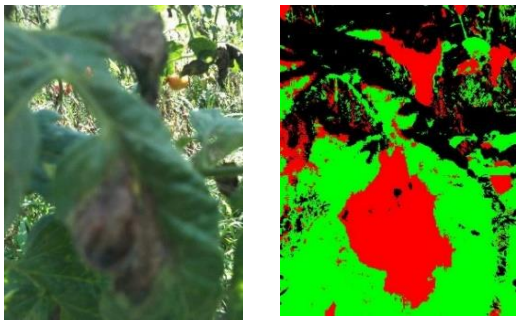


Figura 4. Exemplos de imagens de tomateiros contaminados por *P. infestans* antes e depois a filtragem vermelho/verde, extraído de Vianna e Cruz (2013, 2014).

A identificação de áreas foliares atingidas pelas requeima utiliza imagens reais coletadas no campo experimental sob iluminação solar direta. A filtragem da imagem é um procedimento que considera o princípio de que a cor da planta saudável é a verde intenso (Figura 4). Em seguida é realizada a análise do tom de cada pixel da imagem, classificando-o dentro de uma das seguintes opções: (i) um pixel é classificado como doente e será convertido para a cor vermelha (255,0,0), no sistema RGB, caso seja de um tom amarelado ou marrom, indicativos de algum tipo de lesão. (ii) um pixel é classificado como sadio e será convertido para a cor verde (0,255,0) caso seja de um tom esverdeado, podendo ir de uma matiz mais clara até um verde mais escuro; (iii) um pixel de qualquer tom diferente dos anteriores será considerado como fundo da imagem e será convertido para preto.

Os experimentos realizados por Vianna e Cruz (2013, 2014) com as imagens coletadas e pelo subsistema de redes neurais utilizaram contagens de pixels realizadas sobre as imagens digitais. Em função dessa análise, foram pré-selecionadas para o treinamento das redes neurais oito variáveis, onde cada uma contém informações sobre quantidades de pixels em uma determinada faixa de cor ou tom de cinza. A classificação final das amostras processadas por uma rede neuronal MLP, com configurações variadas, alcançaram um desempenho médio de 90% de acertos (a descrição detalhada está presente nos trabalhos supracitados). Porém, no momento, ainda estão em andamento análises mais detalhadas de desempenho e calibração dos algoritmos de treinamento e teste das redes neuronais.

6. Trabalhos Relacionados

Nosso levantamento bibliográfico não foi capaz de localizar sistemas móveis que apoiam a detecção precoce de requeima em tomateiros. No entanto, do ponto de vista da detecção inteligente já existem trabalhos relacionados. Por exemplo, Sanyal e Patel (2008) conduzirão um estudo de análise de textura e cor das folhas de arroz (*Oryza sp.*) para identificação dados doenças *brown spots* e *blast diseases*. A simulação realizada pelo autor baseou-se em 400 amostras e a classificação foi realizada por uma rede MLP. Foram utilizadas imagens de folhas doentes e normais e chegou-se a um desempenho de 89,26% de acertos na classificação individual dos pixels das imagens. Camargo (2009) define um classificador baseado em uma máquina de vetor de suporte (SVM) para operar sobre as imagens coloridas de algodão

(*Gossypium sp*.). Foram utilizadas 127 imagens de culturas de algodão. O melhor modelo de classificação foi encontrado com 45 características, chegando a 93,1% de desempenho.

Vieira (2011) utilizou uma rede neural MLP para detectar as lesões ocasionadas em folhas de tomate. No entanto, suas taxas de acertos foram de apenas 77,8% na identificação de lesões por requeima. Além disso, dentre os testes conduzidos, foram utilizadas poucas amostras, apenas 9 amostras de folhas com lesões da doença de um conjunto total de 65 amostras de folhas lesionadas por três tipos diferentes de doenças. Neste estudo, existe um fator limitante quando comparado a nossa abordagem, isto é, as imagens digitais sofreram pré-processamento manual, onde o usuário delimita as manchas mais relevantes da imagem original. Além disso, não há informações sobre o tipo de cultivo associados aos tomateiros nem o processo de aquisição das imagens.

7. Conclusão

Devido à globalização da economia e a crescente necessidade de produtos agrícolas, surgiu a necessidade de se obter produções com níveis de qualidade e quantidade cada vez maiores, isso impõe à atividade agrícola novos métodos e técnicas. Concomitantemente, os mercados consumidores estão cada vez mais exigentes com relação à segurança alimentar, rastreabilidade, respeito ao meio ambiente e a saúde do trabalhador rural, barreiras sanitárias e fitossanitárias. Cada vez os produtores percebem que a tomada de decisões e uso de sistemas inteligentes é mais que uma tendência, torna-se uma questão de sobrevivência e necessidade.

Este trabalho apresentou uma abordagem computacional, em fase de finalização de desenvolvimento, mas que já apresenta resultados e ferramentas que poderão ser utilizados na agricultura familiar para detectar precocemente a requeima em tomateiros do estado do Rio de Janeiro e possivelmente aumentar a produtividade da lavoura de tomate e gerar produtos de maior valor agregado.

A principal vantagem da abordagem é contribuir para contribuir para a segurança alimentar e oferecer oportunidades para os pequenos proprietários para executar suas operações de forma mais produtiva baseado em TICs e técnicas básicas de AP. Os artefatos foram projetados para utilizar equipamentos de baixo custo e garantir a facilidade de uso, pois os pequenos agricultores familiares têm pouco espaço para erros e pouca experiência em processamento digital com programas sofisticados ou mesmo o uso de técnicas de redes neurais e bancos de dados. Como trabalhos futuros serão avaliados melhorias ma integração entre os módulos do sistema com os algoritmos de processamento digital de imagens, uma vez que a qualidade dos dados de entrada é o fator de maior impacto para o desempenho de um sistema de reconhecimento baseado em RNA's.

Agradecimentos e Auxílio Financeiro

Agradecemos à FAPERJ pelo financiamento do Projeto (E-26/112.588/2012), ao CNPq e ao FNDE/MEC pelas bolsas concedidas. Agradecimentos ao Instituto de Agronomia da UFRRJ (IA/UFRRJ) e aos profs. Antônio Carlos de S. Abboud e Margarida Goréte F. do Carmo.

7. Referências Bibliográficas

Camargo, A. e Smith, J.S., (2009) "Image Pattern Classification for the Identification of Disease Causing Agents in Plants", Computers and Eletronics in Agriculture 66, p.121-125.

CGI - Comitê Gestor Da Internet No Brasil (2011) "TICs domicílios e empresas 2010". São Paulo, 2011. Disponível em http://www.cetic.br/tic/2010/index.htm

Correa, F.M., Bueno Filho, J.S.S., Carmo, M.G.F (2009)"Comparação de Três diagramáticas Chaves para a quantificação da requeima em folhas de tomate". Plant Pathology: vol. 58 pp 1128-1133.

Coelho, J. P. C. e Silva, J. R. M. (2011) "Agricultura de Precisão". Disponível em http://agrinov.ajap.pt/manuais/Manual_Agricultura_de_Precisao.pdf.

DISQUAL (2010), "Otimização da qualidade e redução de custos na cadeia de distribuição de produtos hortofrutícolas frescos – Manual de Boas Práticas". http://www2.esb.ucp.pt/twt/disqual/pdfs/disqual_tomate.pdf.

Filgueira, F. A. R. (2008) "O novo manual de olericultura", Editora da UFV. 3a edição.

IBGE (2013) – "Brasil em números". Disponível em: Http://biblioteca.ibge.gov.br/visualizacao/ periodicos/2/bn_2013_v21.pdf

MAPA (2013) Ministério da Agricultura, Pecuária e Abastecimento. "Agricultura de precisão" Secretaria de Desenvolvimento Agropecuário e Cooperativismo. – Brasília. ". http://www.agricultura.gov.br/arq_editor/Boletim%20tecnico.pdf.

Nakano, O. (1999) "Como Pragas das hortaliças: Seu Controle e o selo verde". Horticultura Brasileira, vol. 17, n.1 UnB.

Nunes, D., Werly C., Vianna, G.K.Cruz, SMS. (2014) "Early Discovery of Tomato Foliage Diseases Based on Data Provenance and Pattern Recognition". Proc. Of the 5th IPAW. Cologne - Germany.

Quincozes, E. R. F. et al (2010), "Gestão Do Conhecimento Aplicada A Uma Organização Intensiva Em Conhecimento: O Caso Da Embrapa Clima Temperado", Interscience Place, n. 3 p.123-141.

Sanyal, P.; Patel, S. C. (2008), "Pattern recognition method to detect two diseases in rice plants" Imaging Science Journal, v.56, n.6, p. 319-325.

Schwartz, C (2012) "Relações de Gênero e Apropriação de Tecnologias de Informação e Comunicação na Agricultura Familiar de Santa Maria-RS". Tese de Doutorado – UFSM.

Wang, N., et al. (2006) "Wireless sensors in agriculture and food industry-Recente development and Future perpective". Computer and Eletronics in Agriculture, n.50, p.1-14.

Vianna, G.K. e Cruz, S.M.S. (2013) "Análise Inteligente de Imagens Digitais no Monitoramento da Requeima dos Tomateiros". Anais do IX Congresso Brasileiro de Agroinformática. Cuiabá, MT.

Vianna, G.K. e Cruz, S.M.S. (2014) "Using Multilayer Perceptron Networks in Early Detection of Late Blight Disease in Tomato Leaves". Proc. of the 16th annual conference ICAI. Las Vegas, USA.

Vieira, F. S. et. al. (2011) "Utilização de Processamento Digital de Imagens e Redes Neurais Artificiais para Diagnosticar Doenças Fúngicas na Cultura do Tomate". Anais do XX Seminário de Computação, p. 58-69.

MOVE!
Um Sistema de Recomendação de Músicas para uma Atividade Física.

Rodrigo B. Silva, Flavia C. Bernardini

[1]Laboratório de Inovação no Desenvolvimento de Sistemas — LabIDeS
Instituto de Ciência e Tecnologia – Universidade Federal Fluminense (UFF)
Rio das Ostras – RJ – Brazil

{rodrigosilva,fcbernardini}@id.uff.br

Abstract. *Nowadays, it is quite common for people to assemble your playlists for different types of contexts in their daily tasks. This work aims to present an application on Android platform to recommend songs, according to the user's speed, via mobile device. The purpose of this application is to motivate people to engage in physical activity.*

Resumo. *Nos dias de hoje, é bastante comum as pessoas montarem suas listas de músicas para diversos tipos de contextos em seu dia-a-dia. Este trabalho tem como objetivo apresentar um aplicativo em plataforma Android para recomendar músicas, de acordo com a velocidade do usuário, através de um dispositivo móvel. O propósito do aplicativo é motivar pessoas a realizar atividades físicas.*

1. Introdução

A prática de atividade física faz bem ao corpo e à mente. Com ela podemos listar diversos tipos de benefícios que ajudam a lidar com o nosso dia-a-dia, tais como: emagrecer com saúde, diminuir o estresse, melhorar a auto estima, e muitas outras. A atividade física não está restrita a nenhuma idade, podendo assim estar presente desde nossos primeiros anos de vida até a terceira idade. Não existe uma atividade física que se sustente se não houver prazer e satisfação. Com isso é necessária uma motivação que leve as pessoas a realizá-las, que as ajudem a atingir o objetivo e, assim, melhorar a sua saúde. A motivação é uma força interior que se modifica a cada momento durante toda a vida, e direciona e intensifica os objetivos de um indivíduo [Cabral 2014].

Uma das muitas motivações que a pessoas possuem é a realização da atividade física acompanhada por música. A música adequada dá ritmo ao movimento, amplitude e leveza ao corpo. As vibrações musicais provocam vibrações corporais. A música tonifica, exalta, alivia. A música faz com que o participante esqueça um pouco o corpo e as suas fraquezas, com que se purifique pela beleza um gesto em particular, participando ao máximo da atividade [Pavlovic 1987].

A atividade física acompanhada por música ocorre com muita freqüência, seja em situação prática individual, por meio da utilização de fones de ouvidos; seja em situação em grupo, com música ambiente. Em ambas situações, os movimentos executados pelos praticantes podem estar sincronizados com a música, ou esta funcionar

simplesmente como fundo musical. Não se pode negar, entretanto, que muitos a consideram como uma forma de prevenção contra monotonia existente na atividade física sistematizada [Miranda and Godeli 2002].

É comum as pessoas montarem suas listas de músicas para diversos tipos de contextos em seu dia-a-dia, sejam eles festas, reunião de amigos e até mesmo para relaxar, estudar, dentre outros. Ao frequentar uma academia, as pessoas costumam seguir o ritmo da música em execução com a ajuda do professor que prepara as músicas para aquela aula. As pessoas que frenquentam a aula se sentem felizes e motivadas, independentemente se gostam do gênero musical ou do artista. Também é comum observar nas academias que as músicas são selecionadas durante a aula, pois toda vez que uma música termina, o professor sai da sua posição e executa uma outra música manualmente em seu dispositivo móvel, perdendo assim tempo da aula e o foco dos alunos. Assim, automatizar essa montagem de listas de músicas para serem executadas de acordo com uma atividade física pode ser interessante.

O objetivo deste trabalho é propor um aplicativo na plataforma Android para geração de uma lista de músicas recomendadas, utilizando aprendizado de máquina supervisionado, de acordo com a atividade física que o usuário está realizando.

2. Trabalhos Relacionados

Ringo [Maes and Shardanand 1995] é um sistema desenvolvido para recomendação personalizada de música. O trabalho explora similaridades entre os gostos de diferentes usuários para recomendar itens, baseado no fato de que os gostos das pessoas apresentam tendências gerais e padrões entre gostos, e entre grupos de pessoas. As pessoas descrevem suas preferências musicais, avaliando algumas canções. Tais avaliações constituem o perfil dos indivíduos. O sistema usa então esses perfis para gerar recomendações para usuários individuais. Para o seu funcionamento, primeiramente usuários similares são identificados. A partir dessa identificação e comparação de perfis, o sistema pode predizer o quanto o usuário gostaria de um álbum/artista que ainda não foi avaliado pelo mesmo.

Em outro domínio relacionado à recomendação de músicas, muitas vezes a maneira em que uma pessoa dirige está diretamente ligada ao tipo de música que está tocando. Uma música mais suave pode fazer um usuário dirigir mais tranquilamente, e vice-versa. O aplicativo da Volkswagen [Tudo Celular 2014], ao invés de recomendar uma música existente, produz uma música em tempo real, baseando-se na forma como o usuário está dirigindo. O aplicativo funciona coletando informações de um *smartphone* ligado ao carro, usando a velocidade, RPM, aceleração, e informações do GPS.

Runners Rhythm [Runners Rhythm 2014] permite ao usuário selecionar e executar um música nas batidas por minuto (BPM) desejadas para correr ou caminhar. Uma vez que o ritmo for selecionado, ele pode ser ajustado ao ritmo do usuário. No menu de opções é possível ajustar o número de músicas selecionadas por consulta. Também pode ser ajustado o nível máximo ou mínimo de BPM das músicas, o que permite que mais ou menos variabilidade em BPMs quando as músicas são selecionadas durante a atividade.

O sistema apresentado neste trabalho possui como principal diferencial em relação a esses trabalhos apresentados por utilizar aprendizado de máquina para recomendar músicas para os movimentos.

3. Sistemas de Recomendação

Os Sistemas de Recomendação auxiliam no aumento da capacidade e eficácia do processo de indicação de um ou mais itens dentre diversas alternativas, já bastante conhecido na relação social entre os seres humanos. Em um sistema típico, as pessoas fornecem recomendações como entradas que o sistema agrega, e direciona para os indivíduos considerados potenciais interessados neste tipo de recomendação. Um dos grandes desafios desse tipo de sistema é realizar a combinação adequada entre as expectativas dos usuários em relação aos produtos, serviços e pessoas a serem recomendados. Em outras palavras, definir e descobrir esse relacionamento de interesses é o grande problema a ser atacado [Resnick and Varian 1997].

As técnicas de recomendação, em geral, dependem das contribuições dos indivíduos na avaliação da informação. O princípio dos sistemas de recomendação se baseia em "o que é relevante para mim, também pode ser relevante para alguém com interesse similar". Várias técnicas têm surgido visando à identificação de padrões de comportamento (consumo, pesquisa e outros) e utilização de tais padrões na personalização do relacionamento com os usuários, sendo algumas delas: Filtragem Baseada em Conteúdo, Filtragem Colaborativa, Filtragem Hibrida e Filtragem Baseada em Conhecimento [Cazella et al. 2010].

A Filtragem Baseada em Conteúdo tem como objetivo básico utilizar as descrições dos conteúdos dos itens, e comparar com os interesses dos usuários, verificando se o item é ou não relevante para cada um. Essa técnica realiza uma seleção baseada na análise de conteúdo de itens e no perfil do usuário [Herlocker 2000]. As informações sobre o perfil do usuário, onde contém suas preferências e necessidades, podem ser obtidas pelo próprio usuário, como uma consulta realizada por ele, coletadas através do conteúdo dos itens que o usuário consome, ou através de um cadastro de informações [Cazella et al. 2010]. Segundo [Adomavicius and Tuzhilin 2005], a abordagem baseada em conteúdo tem algumas limitações, como por exemplo a extração e análise de conteúdo multimídia (vídeo e som), que é muita mais complexa do que a extração e análise de documentos textuais.

A Filtragem Colaborativa se diferencia da filtragem baseada em conteúdo por não exigir a compreensão ou reconhecimento do conteúdo dos itens. Essa abordagem foi desenvolvida para atender as limitações da filtragem baseada em conteúdo [Herlocker 2000]. Na Filtragem Colaborativa, a essência está na troca de experiências entre as pessoas que possuem interesses comuns. Nesses sistemas, os itens são filtrados baseados nas avaliações feitas pelos usuários. Um usuário de um sistema colaborativo deve, portanto, pontuar cada item experimentado, indicando o quanto esse item casa com sua necessidade de informação. Essas pontuações são coletadas para grupos de pessoas, permitindo que cada usuário se beneficie das pontuações (experiências) apresentadas por outros usuários na comunidade. A filtragem colaborativa pode apresentar algumas limitações, como por exemplo: quando um novo item aparece no banco de dados não existe maneira deste ser recomendado para o usuário até que mais informações sejam obtidas através de outros usuários.

A Filtragem Híbrida busca combinar os pontos fortes da filtragem colaborativa e da filtragem baseada em conteúdo, visando criar um sistema que possa melhor atender as necessidades do usuário [Herlocker 2000]. Essa abordagem é constituída das seguintes vantagens: descoberta de novos relacionamentos entre usuários; recomendação de itens

diretamente relacionados ao histórico; bons resultados para usuários incomuns; e precisão independente do número de usuários.

Por fim, na Filtragem Baseada em Conhecimento, o conhecimento resultante da aplicação do processo de Descoberta de Conhecimento de Base de Dados, ou KDD (do inglês *Knowledge Discovery from Databases*) é utilizado. Segundo [Zaiane 2000], KDD é definido como um processo de extração não trivial de informações potencialmente úteis, as quais não são previamente conhecidas e encontram-se implícitas em grandes coleções de dados. Quando se trabalha com sistemas de recomendação, o KDD se torna um recurso importante para a descoberta de relações entre itens, entre usuários e entre itens e usuários. Através de análises de arquivos de *log*, por exemplo, pode-se obter conhecimentos aprofundados a respeito dos usuários que se conectaram a um website. Este conhecimento pode ser utilizado para a personalização da oferta de produtos, na estruturação de sites de acordo com o perfil de cada internauta e personalização do conteúdo das páginas. Neste trabalho, usamos a técnica de Filtragem Baseada em Conhecimento.

4. Aprendizado de Máquina

Aprendizado de Máquina é uma parte da Inteligência Artificial responsável pelo desenvolvimento de teorias computacionais focadas na criação do conhecimento artificial. Softwares desenvolvidos com esta tecnologia possuem a característica de tomarem decisões com base no conhecimento prévio acumulado através da interação com o ambiente. [Facelli et al. 2011].

Algoritmos de aprendizado têm sido amplamente utilizados em diversas tarefas, que podem ser organizadas de acordo com diferentes critérios. Em uma tarefa de previsão, a meta é encontrar uma função (também chamada de modelo ou hipótese) a partir dos dados de treinamento que possa ser utilizada para prever um rótulo ou valor que caracterize um novo exemplo, com base nos valores de seus atributos de entrada. Para isso, cada objeto do conjunto de treinamento deve possuir atributos de entrada e saída. Os algoritmos ou métodos de Aprendizado de Máquina utilizados nessa tarefa induzem modelos preditivos. Esses algoritmos seguem o paradigma de aprendizado supervisionado. O termo supervisionado vem da presença de um "supervisor externo", que fornece a saída (rótulo) desejada para cada exemplo (conjunto de valores para os atributos de entrada).

No aprendizado supervisionado, o objetivo é induzir um conceito utilizando os exemplos previamente rotulados, denominado conjunto de exemplos de treinamento, realizando generalizações, tal que o conceito (hipótese) induzido é capaz de predizer a classe (rótulo) de futuros exemplos. Nesse caso, o conceito induzido é visto como um classificador [Facelli et al. 2011]. Há várias paradigmas para indução do classificador, e um dele é o simbólico, que envolve os algoritmos de indução de árvores de decisão e de regras. Um algoritmo em aprendizado supervisionado de indução de árvores de decisão é o algortimo J48, uma implementação em Java na ferramenta Weka do algoritmo C4.5 [Witten and Frank 2005]. O modelo de árvore de decisão é construído a partir do processamento dos dados de treino, e o modelo é utilizado para classificar dados ainda não classificados. O J48 gera árvores de decisão, em que cada nó da árvore avalia a existência ou significância de cada atributo individual. As árvores de decisão são construídas do nó raiz para os nós folha, através da escolha do atributo mais apropriado para cada situação. Uma vez escolhido o atributo, os dados de treino são divididos em sub-grupos, correspon-

dendo aos diferentes valores dos atributos, e o processo é repetido para cada sub-grupo até que uma grande parte dos exemplos em cada nó pertençam a uma única classe.

5. Metodologia de Desenvolvimento do Sistema MOVE!

O aplicativo proposto neste trabalho tem como principal objetivo recomendar músicas para uma atividade física. Primeiramente, as músicas da lista do usuário são classificadas por classificador musical. O classificador musical tem como principal objetivo prever, para uma determinada atividade física, qual música é mais adequada. Dada uma música, o classificador recebe dados referentes aos atributos fornecidos pela API da Echo Nest da música, e então a música é classificada para uma atividade física. Para a construção do classificador, foi utilizado o algoritmo de aprendizado J48.

A Echo Nest [1] é a maior empresa de inteligência de música da indústria, fornecendo aos desenvolvedores a compreensão mais profunda do conteúdo de música. Tal conteúdo musical pode ser acessado através da API disponível gratuitamente para o desenvolvimento de aplicativos relacionados a música. A Echo Nest apresenta alguns atributos sonoros utilizados em aplicativos desenvolvidos com a API. Um atributo sonoro é uma qualidade subjetiva estimada em uma faixa de música. Ele é modelado através da aprendizagem e é dado como um único número de ponto flutuante que varia no intervalo fechado $[0, 1]$. A Echo Nest deriva os atributos de uma música das faixas de áudio que a compõem. As canções podem ser classificados por qualquer um desses eixos, ou os atributos podem ser usados como filtros para a construção de listas de reprodução personalizadas.

O conjunto de dados utilizado na construção do classificador foi criado a partir de 390 arquivos em MP3 de diversas músicas de diversos gêneros, estilos e épocas de lançamento. Boa parte das músicas selecionadas para serem analisadas tiveram origem de uma pesquisa de opinião feita através de uma rede social, em que as pessoas recomendavam músicas para algumas atividades.

Os arquivos foram divididos igualmente para três atividades, que foram utilizadas como classes. As atividades escolhidas foram: CORRER, PEDALAR e CAMINHAR. Para a análise de cada música, foi utilizada a biblioteca jEN, para o acesso à API da ECHO NEST, com o objetivo de coletar informações referentes a cada uma. As informações extraídas para atributos de entrada são: *Danceability, Energy, Speechiness, Liveness, Loudness.* Coletamos essas informações de um conjunto de exemplos referentes a 130 músicas para três classes diferentes — CORRER, PEDALAR e CAMINHAR — o que totalizou 390 músicas. O conjunto de dados foi utilizado como entrada para a Ferramenta WEKA, para ser construido o classificador.

5.1. Avaliação do algoritmo de aprendizado

Para a classificação das músicas foi utilizado o algoritmo de indução de árvore de decisão J48 no Weka. Foram utilizados os seguintes atributos, extraídos das músicas pela API da EchoNest: *Danceability*, *Energy*, *Speechiness*, *Liveness* e *Loudness*. O uso de uma árvore de decisão foi devido ao fato de ser uma estrutura simples para ser utilizada em um dispositivo móvel, o qual geralmente possui menos recursos tecnológicos que um

[1] Informações disponíveis em `http://www.echonest.com/`

computador. A estimativa de erro foi executada em validação cruzada de 10 partições, obtendo-se uma baixa acurácia de 38,72%. Para contornar esse problema, foi utilizado o método estatístico de amostragem com reposição, com a aplicação no Weka, do filtro supervisionado *Resample*, para balanceamento das classes. O classifiicador foi induzido novamente com o algoritmo J48 no Weka, em validação cruzada de 10 partições, obtendo-se uma melhor acurácia de 67,45%. A descoberta do conhecimento pela árvore de decisão gerada pelo algoritmo J48 pode ser feita pela visualização de padrões relevantes através de regras de decisão. A árvore de decisão gerada foi implemetada no protótipo contruído. Na Figura 1 é ilustrado o método para construção do classificador.

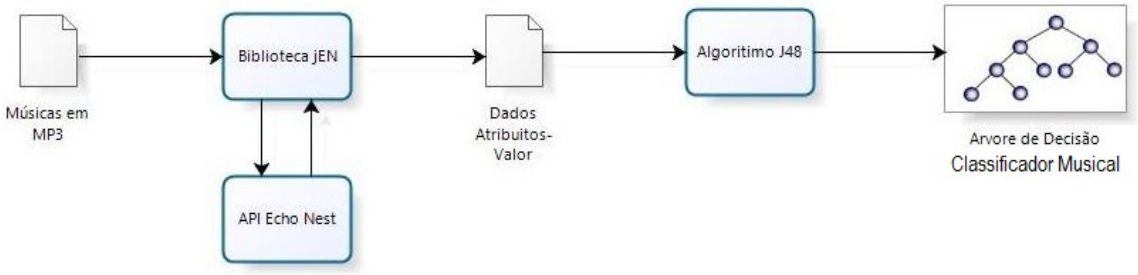


Figura 1. Construção do Classificador

Na Figura 2 é apresentada o processo de classificação de uma música. A música em formato MP3 é dada como entrada para a biblioteca jEN, que fará a análise desta música e extrair seus dados músicais referentes a *Danceability, Energy, Speechiness, Liveness, Loudness* sendo assim passando para a próxima etapa que é o classificador musical onde uma atividade física será atribuída para a música e por final teremos a música classificada. Podemos notar a tarefa de classificação é simples, no intuito de acelerar esse processo, devido as limitações de um dispositivo móvel em termos de hardware e software serem bastantes limitados.

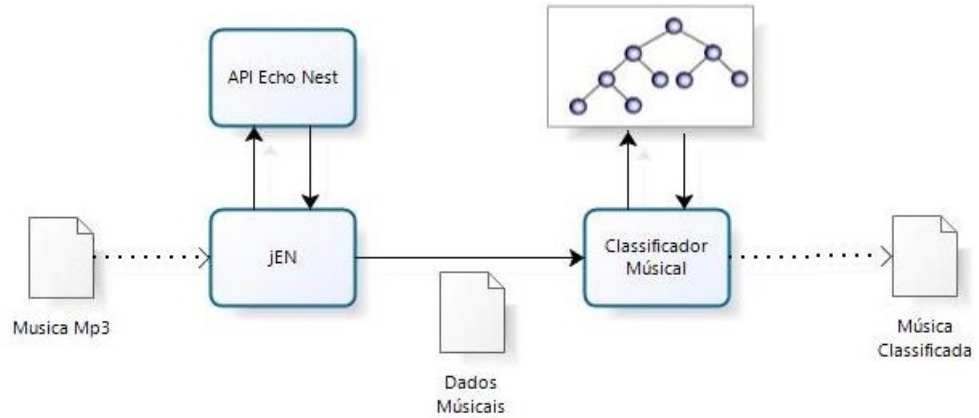


Figura 2. Processo de Classificação

5.2. Componentes do Sistema Move

Na Figura 3 é apresentado o diagrama de componentes do sistema MOVE!. O componente do sistema MOVE! utiliza internamente a plataforma Android. Esse componente se

comunica com o componente da API da Echo Nest, no qual são extraídas as características das músicas na Internet. O componente GPS tem como principal função fornecer a velocidade do usuário, utilizada para identificar o movimento do usuário — ANDAR (de 0 Km/h a 7 Km/h), CORRER (de 7,1 km/h a 14 Km/h), PEDALAR (a partir de 14,1 Km/h). O componente de banco de dados SQlite tem como objetivo armazenar os dados sobre as músicas do aplicativo. Os componentes Musica.java e Playlist.Java são as principais classes do aplicativo. Deve ser observado que na classe Musica.java está implementado o método para classificação musical.

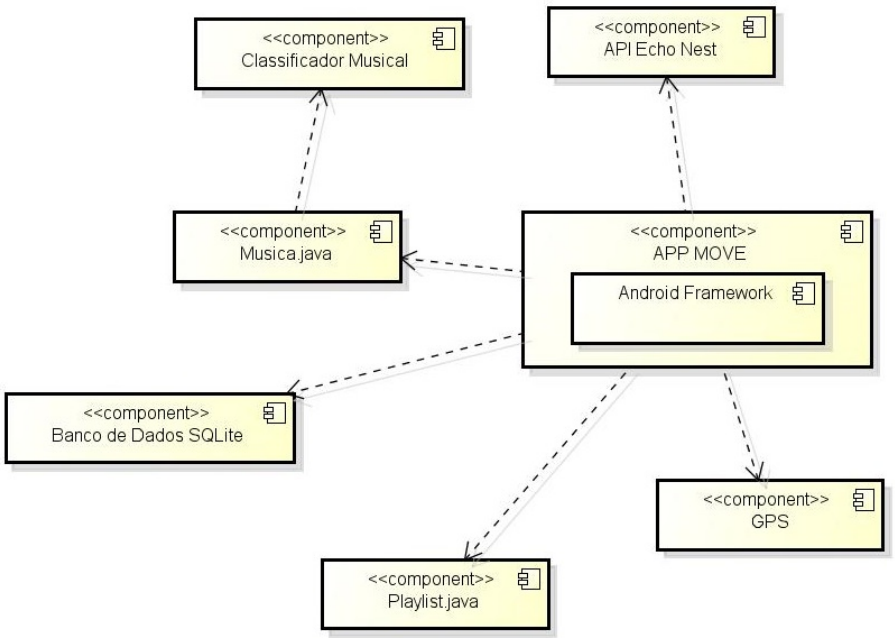


Figura 3. Diagrama de Componentes

6. Conclusão

Neste trabalho, apresentamos um sistema de recomendação de músicas de acordo com a atividade do usuário. O protótipo foi implementado utilizando a API da Echo Nest e a Plataforma Android, que por sua vez utiliza a linguagem Java. Para o aplicativo recomendar músicas, foi utilizado um método que utiliza uma árvore de decisão, induzida utilizando o algoritmo J48, implementado na ferramenta WEKA. Para induzir o classificador, foram utilizadas recomendações oferecidas por usuários de uma rede social. A taxa de erro obtida com o classificador induzido foi considerada promissora, pois o usuário do aplicativo pode ter, durante sua atividade, a melhor lista musical sugerida baseadas em sugestões de outros usuários, e que contém apenas suas músicas favoritas. Porém, como a API necessita de uma conexão com a Internet, a música pode ter seu tempo de análise aumentado caso essa mesma conexão seja lenta. Pelo resultado da análise do conjunto de músicas feita para a indução da árvore de decisão, fica claro que é de extrema importância a pesquisa de nossos atributos para aprimorar ainda mais a classificação da atividade para a qual a música se encaixa.

Observamos, com o desenvolvimento deste trabalho, que dados musicais são um grande desafio. Pela pesquisa em redes sociais, ficou claro que as preferências dos

usuários são bastante heterogêneas. Ao contrário do esperado, observamos que músicas com batidas enérgicas e pesadas podem ser recomendadas para atividades físicas tranquilas, e o inverso também é possível. Neste trabalho, pudemos observar que pode ser mais interessante explorar construção de um classificador para cada indivíduo, ao invés de construir om único classificador para todos os usuários, devido a essa característica intrínseca de cada usuário. Ainda, observamos também que definir o movimento de cada usuário também é um desafio que deve ser explorado, já que o movimento ANDAR de uma pessoa pode ser equivalente ao movimento CORRER de outra. Por fim, experimentos com usuários também necessitam ser realizados com nosso protótipo.

Referências

Adomavicius, G. and Tuzhilin, A. (2005). *Toward the Next Generation Of Recomender Systems: A Survey of the State-of-the Art and Possible Extensions.* IEEE Transations on Knowledgeand Data Engineering.

Cabral, G. (2014). Motivação. DisponÃvel em`http://www.brasilescola.com/psicologia/motivacao-psicologica`. Acessado em 23 de maio de 2014].

Cazella, S. C., Nunes, M. A. S. N., and Reategui, E. B. (2010). *A Ciência da Opinião: Estado da arte em Sistemas de Recomendação*. XXX Congresso da SBC: Computação Verde – Desafios Científicos e Tecnológicos.

Facelli, K., Lorena, A. C., Gama, J., and Carvalho, A. (2011). *Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina*. LTC.

Herlocker, J. L. (2000). *Understanding and Improving Automated Collaborative Filtering Systems*. Tese de Doutorado em Ciência da Computação, University of Minnesota.

Maes, P. M. and Shardanand, U. (1995). Social information filtering: algorithms for automating "word of mouth". In *Proc. SIGCHI Conference on Human Factors in Computing Systems*.

Miranda, M. L. J. and Godeli, M. R. C. S. (2002). *Avaliação de Idosos sobre o papel e a influência da música na atividade física*. Dissertação de Mestrado — USP.

Pavlovic, B. (1987). *Ginástica aeróbica – Uma nova cultura física*. Rio de Janeiro: Sprint.

Resnick, P. and Varian, H. R. (1997). *Recommender Systems*. Comm. of the ACM.

Runners Rhythm (2014). Runners rhythm. Disponível em `http://www.runnersrhythm.com/index.html`. Acessado em 21 de Julho de 2014.

Tudo Celular (2014). App da volkswagen cria música baseada em como você dirige. Disponível em `http://www.tudocelular.com/curiosidade/noticias/n31414/appcarro-volkswagen-musica-trilha-sonora.html`. Acessado em 21 de Julho de 2014.

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2nd edition.

Zaiane, O. R. (2000). *Web Mining: Concepts, Practices and Research*. Simpósio Brasileiro de Banco de Dados.

Aplicação de Métodos baseado em Processos de Negócio para Desenvolvimento de Serviços

Luan Lima[1], Ricardo Diniz Sul[1,2], Leonardo Guerreiro Azevedo[1,2,3]

[1] Departamento de Informática Aplicada (DIA)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro – RJ – Brasil

[2] Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro – RJ – Brasil

[3]IBM Research

{luan.lima, ricardo.diniz, azevedo}@uniriotec.br, LGA@br.ibm.com

Abstract. *SOA is fundamental to support organizations in adapting their systems to the constant environment changes. Service development is not an easy task. A systematic method should be used to guide the work. This paper presents a practical application of services development methods for identification, analysis, design and implementation, and the use of an Enterprise Service Bus (ESB) to support the service deployment. Oracle Service Bus was chosen after evaluating commercial and open source tools. Although specifics methods were used for service development, this work brings insights and practical experiences that are helpful for SOA Analysts and SOA Developers to use in their environment.*

Resumo. *SOA é fundamental para as organizações conseguirem se adaptar às constantes mudanças. O desenvolvimento de serviços não é uma tarefa simples. Um processo sistemático deve ser utilizado para guiar o trabalho. Este trabalho apresenta a aplicação prática de métodos para desenvolvimento de serviços para identificação, análise, projeto e implementação, além da utilização de um Enterprise Service Bus (ESB) para apoiar a implantação dos mesmos. O Oracle Service Bus foi escolhido após análise entre ferramentas comerciais e de código aberto. Embora métodos específicos tenham sido utilizados, este trabalho traz novas ideias e experiências práticas úteis para Analistas SOA e Desenvolvedores SOA aplicarem em seus ambientes.*

1. Introdução

Os ambientes organizacionais vem se alterando constantemente, criando oportunidades assim como riscos e ameaças para as organizações. Service-Oriented Architecture (SOA) é uma abordagem para lidar com estas mudanças e aumentar a agilidade e a capacidade de resposta a mudanças (Kohlborn *et al.*, 2009). Um serviço representa uma funcionalidade autocontida correspondente a uma atividade do negócio (Josuttis, 2007).

Organizações possuem inúmeros processos de negócio. Aplicações e serviços são desenvolvidos para dar apoio à execução das atividades destes processos. Um processo para o desenvolvimento de serviços é necessário a fim de evitar problemas, como, por exemplo, serviços duplicados, sem reuso, com alto acoplamento e pouca coesão e que não atendam às necessidades do negócio. Este objetivo pode ser alcançado através do uso de um método para desenvolvimento de serviços a partir de modelos de processo de negócio [Kohlborn *et al.*, 2009]. Um modelo de processos de negócio possui diversas informações detalhadas como, por exemplo, regras de negócio, requisitos de negócio, atividades, informações de entrada e saída e fluxo de atividades a partir das quais serviços podem ser identificados para serem implementados. Azevedo *et al.* (2009a, 2011, 2013) realizaram uma análise da literatura e concluem que falta sistemática nos métodos para desenvolvimento de serviços a partir de modelos de processos de negócio. Eles apresentam apenas diretrizes sem detalhes suficientes para aplicá-los na prática para identificação, análise, projeto e implementação de serviços.

Este trabalho tem o objetivo de aplicar os métodos baseados em processos de negócio para desenvolvimento propostos por Azevedo *et al.* (2009, 2011 e 2013) e Diirr *et al.* (2012). As heurísticas propostas nestes trabalhos foram executadas em um cenário fictício com tamanho suficiente para permitir: (i) Analisar o uso efetivo dos métodos; (ii) Produzir ideias e experiências práticas úteis para Analistas e Desenvolvedores SOA. Além disso, os serviços foram disponibilizados em um *Enterprise Service Bus* (ESB), o que não foi abordado até o momento nos trabalhos estudados. ESB é uma das principais tecnologias para a implantação de SOA (Hewitt, 2009). O Oracle Service Bus (OSB) foi escolhido após análise das ferramentas disponíveis no mercado. O OSB encontra-se entre as três melhores ferramentas do mercado (Vollmer *et al.*, 2011).

Este trabalho está dividido da seguinte forma. A Seção 1 apresenta a introdução, explicitando o contexto do trabalho. A Seção 2 apresenta os métodos para desenvolvimento de serviços. A Seção 3 apresenta a aplicação dos métodos. Finalmente, a Seção 4 apresenta a conclusão do trabalho.

2. Método para Desenvolvimento de Serviços

Esta seção apresenta os métodos empregados neste trabalho.

2.1. Identificação

A identificação de serviços candidatos possui como entrada modelos de processos de negócio e gera como saída um conjunto de serviços candidatos com informações consolidadas dos mesmos (Azevedo *et al*. 2009). O método de identificação é dividido em três passos: selecionar atividades do modelo de processos de negócio; identificar e classificar os serviços candidatos; e consolidar os serviços candidatos.

O primeiro passo seleciona as atividades que são totalmente executadas por sistemas, por pessoas com auxílio de sistemas, ou candidatas para automação. No segundo passo, serviços são identificados com o uso de heurísticas que analisam elementos semânticos e estruturais dos fluxos do processo, identificando e classificando os serviços como candidatos de dados (i.e., realizam operações CRUD – Create, Retrieve, Update e Delete sobre dados) ou de lógica (i.e., encapsulam regras de negócio). O terceiro passo executa heurísticas para consolidar as informações extraídas

dos serviços em grafos, gráficos e tabelas para auxiliar o analista SOA a decidir quais serviços candidatos serão implementados como serviços físicos e quais serão implementados como funcionalidades de aplicações. Ao final, tem-se uma lista de serviços candidatos e informações que serão utilizadas nos passos seguintes.

2.2. Análise

A análise de serviços candidatos (Azevedo *et al*., 2011, 2013) é dividida em três passos: priorizar serviços candidatos; definir a granularidade dos serviços candidatos; e agrupar os serviços candidatos.

O primeiro passo define a prioridade dos serviços candidatos a partir da soma de valores (pesos) estabelecidos para cada informação relevante do serviço, tais como, por exemplo, grau de reuso e associação com sistemas. No segundo passo, mapas de granularidade são elaborados a fim de separar serviços de granularidade fina e de granularidade grossa, além de explicitar dependências entre eles. O terceiro passo consiste em agrupar os serviços candidatos de acordo com as entidades que estes manipulam. Neste passo, mapas de granularidade são utilizados para auxiliar no agrupamento dos serviços. Ao final, têm-se grupos de serviços candidatos de acordo com as normas da organização.

2.3. Projeto

O projeto de serviços (Diirr *et al.*, 2012) é dividido em quatro passos: separar serviços de dados de serviços de lógica; elaborar o modelo canônico das entidades manipuladas pelos serviços; definir as operações dos serviços; e modelar os serviços.

O primeiro passo considera que serviços de dados e serviços de lógica possuem diferentes características e, portanto, devem ser implementados de formas diferentes. No segundo passo, o modelo canônico deve ser elaborado, a fim de reduzir transformações de dados na comunicação entre serviços. O terceiro passo define como será a implementação dos serviços candidatos: (i) Implementado como uma operação de serviço físico; (ii) Um ou mais serviços candidatos implementados em uma única operação de serviço físico; (iii) Decomposto em mais de uma operação de serviço físico. O quarto passo realiza a modelagem dos serviços utilizando UML. Ao final, tem-se projetado como serviços e suas operações deverão ser implementados.

2.4. Implementação

A etapa de implementação (Diirr *et al.*, 2012) é dividida em dois passos: implementar entidades do negócio; e, implementar serviços.

No primeiro passo, devem ser implementadas as classes para as entidades do modelo canônico. Estas entidades representam os tipos de dados manipulados pelo negócio e, em geral, estarão ligadas diretamente aos dados armazenados no banco de dados. No segundo passo, serviços são implementados utilizando tecnologia de web services. Erl (2005) aponta que esta tecnologia é a principal tecnologia para implementação de serviços. Ao final, tem-se os serviços implementados.

3. Aplicação do Método

O cenário utilizado para a aplicação do método corresponde à compra e venda de imóveis pela internet, considerando a parceria de diversas imobiliárias com uma imobiliária fictícia chamada de “Imobiliária Online” em um portal de vendas, onde são disponibilizados os imóveis de todas as envolvidas para os clientes. A partir deste cenário foi elaborado o diagrama de macroprocesso “Comprar imóvel” apresentado na Figura 1. Este trabalho foca no processo “Submeter proposta de compras” cujos detalhes são apresentados na Figura 2.

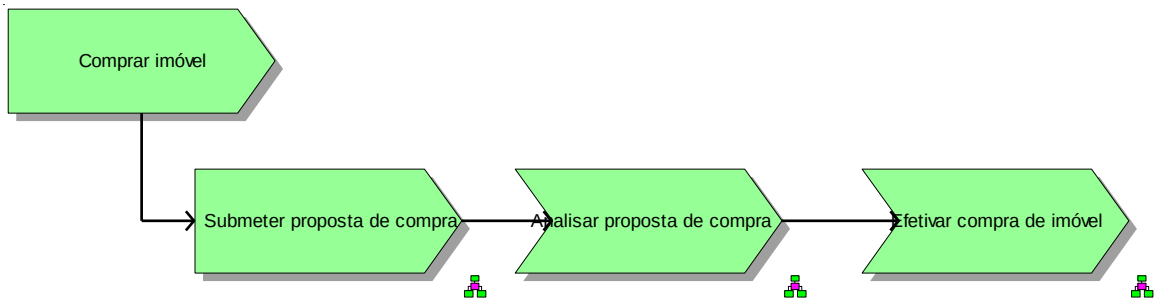


Figura 1 - Diagrama do macro processo "Comprar imóvel"

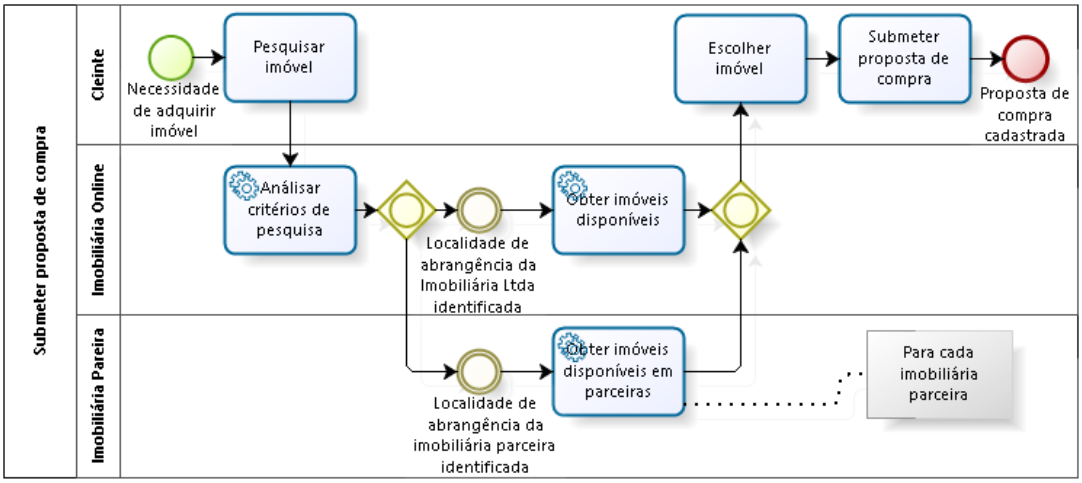


Figura 2 - Processo "Submeter proposta de compra"

Este processo é um processo novo, e, portanto, tanto a Imobiliária Online como suas parceiras não possuem nenhuma implementação. SOA é uma abordagem adequada neste caso para a integração entre as imobiliárias, uma vez que estas se encontram em locais físicos distintos e podem ter seus sistemas implementados em linguagens diferentes. Além disso, caso se inicie uma nova parceria, SOA levaria essa integração a acontecer de forma rápida e pouco custosa. A partir destas considerações, o cenário técnico para integração das imobiliárias foi elaborado e é apresentado na Figura 3.

Para facilitar a integração entre as aplicações das imobiliárias foi utilizado um Enterprise Service Bus (ESB) para implantar os serviços. Para tal, foi desenvolvido um portal disponibilizando uma interface gráfica para os clientes, permitindo acesso aos serviços. Cada imobiliária possui seus imóveis armazenados em um banco dados próprio, com estruturas distintas, o que foi simulado com o desenvolvimento dos bancos

em diferentes idiomas. A seguir são apresentados os resultados da aplicação dos métodos de desenvolvimento de serviços.

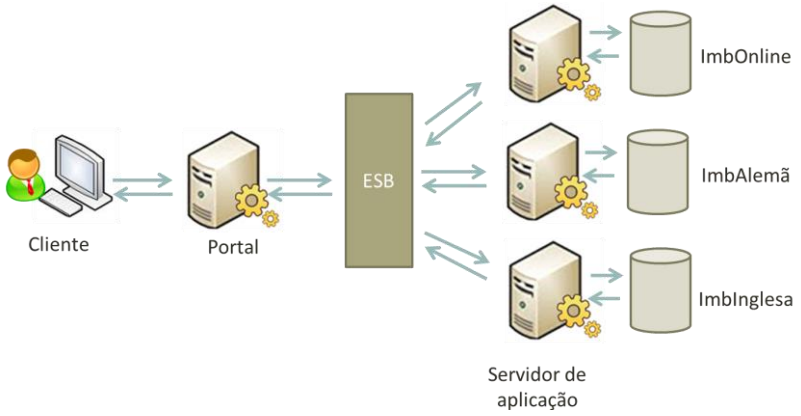


Figura 3 - Arquitetura utilizada

3.1. Identificação

A partir da aplicação das heurísticas de identificação, os dez serviços identificados resultantes são apresentados na Tabela 1. Maiores detalhes do passo a passo da aplicação das heurísticas são apresentados por Lima e Sul (2012).

Tabela 1 - Listagem dos serviços candidatos

Serviços candidatos	
1. Registrar critérios de pesquisa	6. Invocar serviços de imobiliárias parceiras
2. Verificar abrangência de imobiliárias parceiras	7. Consolidar resultado da pesquisa de imóveis
3. Listar imóveis da imobiliária Online	8. Reservar imóvel
4. Listar imóveis de imobiliárias parceiras	9. Registrar proposta de compra
5. Buscar imóveis	10. Submeter proposta de compra

3.2. Análise

Para realização da análise de serviços, o modelo canônico foi elaborado e grupos de entidades foram identificados, como mostrado na Figura 4.

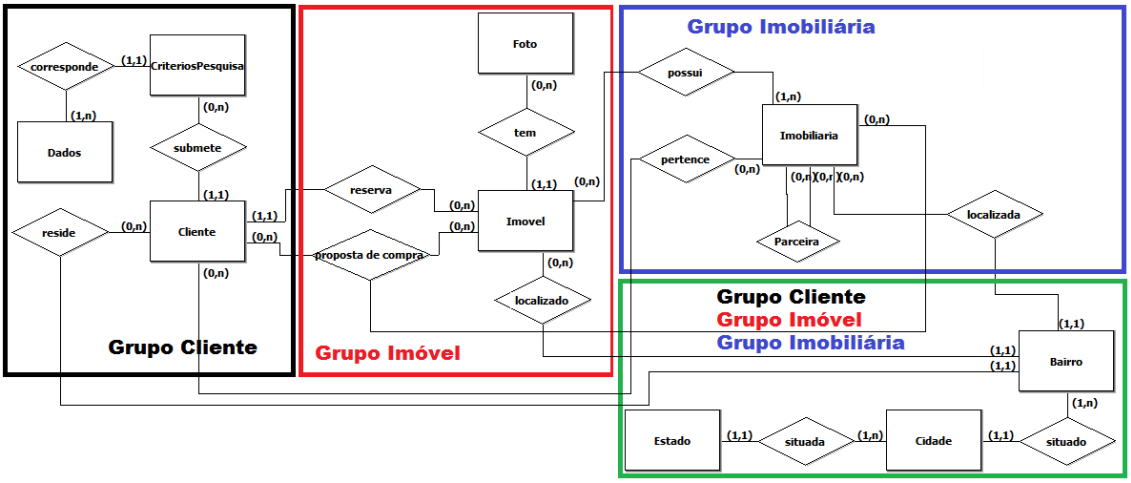


Figura 4 - Grupos de entidades

3.3. Projeto

Esta seção apresenta o projeto dos serviços candidatos que serão implementados. A etapa de projeto se inicia com a separação dos serviços por tipo (dados ou lógica), como pode ser observado na Tabela 2. Vale ressaltar que o serviço *Registrar proposta de compra* e *Submeter proposta de compra* (Tabela 1) foram unificados no serviço *Registrar proposta de compra*, a partir da heurística de seleção de serviços.

Tabela 2 - Identificação dos tipos de serviços

Serviços candidatos	Tipo do serviço
Registrar critérios de pesquisa	Dados
Verificar abrangência de imobiliárias parceiras	Dados
Listar imóveis da imobiliária Online	Dados
Listar imóveis de imobiliárias parceiras	Dados
Buscar imóveis	Dados
Invocar serviços de imobiliárias parceiras	Dados
Consolidar resultado da pesquisa de imóveis	Dados
Reservar imóvel	Dados
Registrar proposta de compra	Dados

3.4. Implementação

Nesta etapa, os serviços listados na Tabela 3 foram implementados utilizando a ferramenta NetBeans IDE e, em seguida, foram disponibilizados no Oracle Service Bus.

De acordo com as necessidades de implementação, alguns serviços foram implementados na aplicação cliente, como os serviços *Buscar imóveis*, *Consolidar resultado de pesquisa de imóveis* e *Invocar serviços de imobiliárias parceiras*. Isto ocorreu porque a orquestração dos serviços foi realizada pela aplicação cliente.

O serviço *Registrar proposta de compra* foi considerado como um serviço de lógica (Tabela 3), uma vez que é verificado se o imóvel possui ou não uma proposta de compra aprovada. No caso de possuir uma proposta de compra aprovada, o sistema deve retornar ao cliente que o imóvel já foi vendido e, caso contrário, deve ser feito um registro da proposta de compra que corresponde à submissão da mesma para análise (processo "*Analisar proposta de compra*"). Este processo está fora do escopo deste trabalho. O serviço *Reservar imóvel* não foi implementado uma vez que durante discussões para desenvolvimento da aplicação, identificamos que o imóvel só está reservado a um cliente se a proposta de compra for aceita pela imobiliária proprietária do imóvel. Logo, a regra de negócio "Reservar imóvel", a partir da qual derivou o serviço, foi revista na especificação do processo de negócio.

Os serviços *Listar imóveis de imobiliárias parceiras* e *Verificar abrangência de imobiliárias parceiras* foi implementado para cada imobiliária parceira, resultando nos serviços: *Listar imóveis de imobiliária alemã* e *Listar imóveis de imobiliária inglesa,* além de *Verificar abrangência de imobiliária alemã* e *Verificar abrangência de imobiliária inglesa.* Os serviços *Listar imóveis da imobiliária Online, Listar imóveis de imobiliária alemã* e *Listar imóveis de imobiliária inglesa*, possuem a mesma saída que são todos os dados referentes ao imóvel e o CNPJ da imobiliária.

Já os serviços *Buscar imóveis*, *Consolidar resultado de pesquisa de imóveis* e *Invocar serviços de imobiliárias parceiras* foram implementados na aplicação cliente, e

consolidados como o serviço *Consolidar busca* (Tabela 3). Este serviço tem como objetivo invocar os serviços *Listar imóveis da imobiliária Online*, *Listar imóveis de imobiliária alemã* e *Listar imóveis de imobiliária inglesa* de acordo com as imobiliárias que atendem aos bairros pesquisados e consolidar em uma lista única os seus retornos.

Outra observação importante é que sempre que uma proposta de compra é registrada (serviço *Registrar proposta de compra*) para um imóvel de uma imobiliária parceira, os dados referentes ao imóvel em questão são armazenados junto ao banco de dados da imobiliária Online, uma vez que é necessário o registro dos dados do imóvel para referência, e caso o imóvel já possua uma proposta de compra "*Aprovada*" o registro da proposta é rejeitado.

Tabela 3 - Serviços implementados

Nome do serviço implementado	Tipo do serviço
Verificar abrangência da imobiliária alemã	Dados
Verificar abrangência da imobiliária inglesa	Dados
Verificar abrangência da imobiliária Online	Dados
Consolidar abrangência	Dados
Registrar critérios de pesquisa	Dados
Listar imóveis da imobiliária alemã	Dados
Listar imóveis da imobiliária inglesa	Dados
Listar imóveis da imobiliária online	Dados
Consolidar busca	Dados
Registrar proposta de compra	Lógica

4. Conclusão

Neste trabalho foi realizada a aplicação prática dos métodos para identificação, análise, projeto e implementação de serviços propostos por Azevedo *et al.* (2009, 2013) e Dirr *et al.* (2012), além de disponibilizar os serviços implementados em um ESB e desenvolver uma aplicação web para consumir estes serviços. O ESB utilizado foi o Oracle Service Bus.

A utilização das heurísticas definidas por Azevedo *et al.* (2009, 2013) e Diirr *et al.* (2012) foi de grande importância para o desenvolvimento dos serviços, pois auxiliaram como um guia em cada etapa de desenvolvimento, além de proporcionar, entre outras vantagens, uma maior facilidade de reuso e um mapeamento alinhado ao negócio. As etapas de identificação, análise e projeto de serviços foram facilitadas pelas explicações e exemplos descritos nos referidos trabalhos, produzindo níveis de detalhes suficientes para agilizar a etapa de implementação.

Durante a aplicação do método, notou-se a carência de informações detalhadas no modelo de processo de negócio utilizado, sendo necessária a realização de ajustes neste modelo, complementando com maiores informações. Com isto, verificou-se que para a aplicação do método em cenários reais é de grande importância que os processos sejam bem detalhados, com a explicitação de todas as regras de negócio associadas ao processo e descrições bem claras, para que se obtenham bons resultados a partir da aplicação do método estudado.

Com base no cenário utilizado, a aplicação de um ESB foi de grande importância pois promoveu a integração entre as imobiliárias, mostrando-se uma ferramenta de grande relevância para projetos com sistemas distribuídos, visto que diminui o

acoplamento, permitindo que aplicações clientes façam referência direta ao ESB sem a necessidade de conhecer a localização dos serviços, o que é interessante caso seja incluída uma nova imobiliária parceira, por exemplo.

Como trabalhos futuros propomos: elaboração de heurísticas para explicitar como o uso de ESB deve ser feito nas fases de projeto e implementação de serviços; elaboração de heurísticas para integração de dados, uma vez que é frequente existir em aplicações distribuídas necessidade de integração de dados com estrutura e semânticas diferentes; uso dos métodos em um cenário real.

Referências

AZEVEDO, L. G.; SANTORO, F.; BAIAO, F.; SOUZA, J. F.; REVOREDO; PEREIRA; HERLAIN (2009). "A Method for Service Identification from Business Process Models in a SOA Approach". In: 10th Workshop on Business Process Modeling, Development, and Support (BMDS'09), v. 29. p. 99-112.

AZEVEDO, L.G.; BAIAO, F.; SANTORO, F.; SOUZA, J. (2011). A Business Aware Service Identification and Analysis Approach. In: International Conference Information Systems (IADIS), Avila, Spain.

AZEVEDO, L. G., SANTORO, F. M., BAIÃO, F., DIIRR, T., SOUZA, A., SOUZA, J. F., and SOUSA, H. P. (2013) "A method for bridging the gap between business process models and services". iSys: Revista Brasileira de Sistemas de Informação, 6(1):62–98.

DIIRR, T.; AZEVEDO, L. G.; SANTORO, F.; BAIAO, F.; FARIA, F. (2012). "Practical Approach for Service Design and Implementation". IADIS International Conference in Information Systems, Berlim, Alemanha.

ERL, T. (2005).Service-Oriented Architecture: concepts, technology, and design. Prentice Hall.

HEWITT, E. (2009). Java SOA Cookbook. O'Reilly.

JOSUTTIS, N. (2007). SOA in practice: The Art of Distributed System Design. O'Reilly.

KOHLBORN, T.; KORTHAUS, A.; CHAN, T.; ROSEMANN, M. (2009). "Identification and Analysis of Business and Software Services - A Consolidated Approach". In: IEEE Transactions on Services Computing.

LIMA, L. F. M. da M.; SUL, R. D. (2012). "Aplicação de Método baseado em Processos de Negócio para Desenvolvimento de Serviços". Monografia de Trabalho de Conclusão de Curso de Graduação. Departamento de Informática Aplicada, Universidade Federal do Estado do Rio de Janeiro (UNIRIO).

MARKS, E. A.; BELL, M. (2006). Service-Oriented Architecture: a planning and implementation guide for business and technology. John Willey & Sons Inc.

VOLLMER, K., GILPIN, M., ROSE, S (2011). The Forrester Wave™: Enterprise service Bus, Q2 2011. The Forrester Research.

Desenvolvimento de web services interoperáveis utilizando abordagem contract-first

Marcos Mele[1,2], Sandro Lopes[1], Leonardo Guerreiro Azevedo[1,2,3]

[1]Departamento de Informática Aplicada
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

[2]Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

[3]IBM Research

{marcos.mele, sandro.lopes, azevedo}@uniriotec.br, LGA@br.ibm.com

Abstract. *Service-Oriented Architecture (SOA) supports interoperability among heterogeneous systems. However, even using SOA, integration between systems is a challenge. In practice, usually the service contract is generated automatically from service code (code-first approach), bringing service maintenance and consumption issues in the long term. A better approach is to build the service contract first and implement the code adhering to this contract (contract-first approach). This work explores contract-first development through an example of use in order to demonstrate this approach brings a high level of interoperability, makes governance ease and reduces coupling with a particular technology.*

Resumo. *A Arquitetura Orientada a Serviços (SOA) apoia a interoperabilidade entre sistemas heterogêneos. Entretanto, mesmo usando SOA, a integração entre sistemas é um desafio. Na prática, o contrato do serviço geralmente é gerado automaticamente a partir do código (abordagem code-first), o que gera dificuldades de manutenção e consumo do serviço em longo prazo. Uma melhor abordagem é construir o contrato do serviço primeiro e, então, implementar o código aderente a este contrato. Este trabalho faz um estudo exploratório do desenvolvimento contract-first através de um exemplo de uso, a fim de demonstrar que esta abordagem traz alto nível de interoperabilidade, facilita governança e reduz acoplamento de tecnologia.*

1. Introdução

Segundo Josuttis [2007], SOA (Service-Oriented Architecture) possibilita a comunicação entre sistemas desenvolvidos em linguagens de programação distintas e sendo executados em sistemas operacionais diferentes. Esse cenário é muito comum em empresas de médio e grande porte que possuem uma infraestrutura de TI complexa.

SOA é um paradigma arquitetural que disponibiliza funcionalidades de aplicações de maneira autocontida como serviços [Erl, 2005]. Serviços são componentes de software que representam uma atividade do negócio [Hewitt 2009]. Em geral, serviços são desenvolvidos utilizando a tecnologia de *web services* e são responsáveis

pela comunicação e interoperabilidade entre sistemas. Usualmente, serviços são descritos utilizando a linguagem WSDL[1] (Web Services Description Language) e a comunicação é feita de forma padronizada utilizando troca de mensagens escritas segundo o protocolo SOAP[2] (Simple Object Access Protocol).

Como em SOA a comunicação entre provedor e consumidor é estabelecida através do contrato do serviço [Josuttis, 2007], cada serviço desenvolvido terá primeiro seu contrato estabelecido, evitando que mudanças e tendências durante o desenvolvimento afetem seus consumidores. Alguns padrões e convenções serão abordados neste trabalho, a fim de garantir interoperabilidade, redução de acoplamento, aumento de reusabilidade e melhoria em governança dos serviços.

O objetivo deste trabalho é apresentar abordagem de como se desenvolver serviços através do uso de *contract-first*. Esta abordagem propõe desenvolver os artefatos relativos ao contrato de serviço antes de sua codificação [Erl, 2008]. A utilização desse modo de desenvolvimento auxilia na criação de serviços com maior nível de independência, eficiência e baixo acoplamento à tecnologia. A proposta deste trabalho para utilização de *contract-first* é baseada no uso de padrões e boas práticas de desenvolvimento de serviços presentes na especificação do Basic Profile (BP)[3] que visam garantir interoperabilidade, redução de impacto de mudanças e facilidade de manutenção de serviços.

O foco deste trabalho está no desenvolvimento do serviço em si. Logo, foram escolhidos os resultados apresentados por Azevedo *et al.* [2013] para as fases anteriores do processo de desenvolvimento (isto é, identificação e análise de serviços).

A abordagem escolhida para implementação de serviços foi web services e a implementação da composição de serviços utilizando injeção direta no serviço composto. Os contratos dos serviços foram desenvolvidos utilizando WSDL, que é linguagem padrão de definição de web services; enquanto a API JAX-WS foi escolhida para desenvolvimento de web services devido a sua robustez e praticidade por permitir uso de *annotations* (padrão de mapeamento por anotação de classes Java).

Este trabalho está dividido da seguinte forma. A Seção 1 é a presente introdução. A Seção 2 apresenta o conceito de contrato de serviço e nossa proposta de projeto de serviços interoperáveis, apresentando o Basic Profile (BP), que corresponde a padrões de desenvolvimento de serviços com maior interoperabilidade, e traz um comparativo das abordagens *contract-first* e *code-first*. A Seção 3 trata do desenvolvimento dos serviços utilizando o framework JAX-WS 2.2 e outras especificações J2EE. Por fim, a seção 4 apresenta análises dos resultados alcançados e a conclusão do trabalho.

2. Contrato de Serviços: projetando web services interoperáveis

Um contrato de serviço determina as operações, o formato dos dados, e os detalhes técnicos de um serviço definindo o acordo entre um provedor específico para um

[1] http://www.w3.org/TR/wsdl

[2] http://www.w3.org/TR/soap/

[3] http://ws-i.org/Profiles/BasicProfile-1.2-2010-11-09.html

consumidor específico [Josuttis, 2007]. Desta forma, um contrato deve conter informações sobre os requisitos do serviço, incluindo aspectos funcionais (operações e dados de entrada e saída) e não-funcionais (como qualidade e segurança). Para serviços baseados em SOAP, o WSDL representa a parte técnica do contrato [Hewitt, 2009]. O método que constrói o WSDL primeiro e posteriormente codifica os requisitos em uma linguagem de programação (como, por exemplo, Java) a partir das definições presentes no WSDL é chamado *contract-first*. Em contrapartida, o método onde primeiro é feita a codificação do serviço e a partir da codificação o WSDL é gerado é chamado *code-first* ou *contract-last*.

2.1. Conformidade com o WS-I Basic Profile

SOA é um paradigma para a realização e manutenção de processos de negócio em um grande ambiente de sistemas distribuídos que são controlados por diferentes proprietários [Josuttis, 2007]. A integração entre sistemas que utilizam distintas tecnologias de desenvolvimento e plataformas é viável através da utilização de serviços independentes de plataforma e protocolo. O padrão de fato para desenvolvimento de serviços é web services [Erl, 2008]. Contudo, para garantir sua utilização pelo maior número de consumidores possíveis, é importante seguir boas práticas de desenvolvimento. Este trabalho considera o padrão web services utilizando o protocolo SOAP cujos principais padrões empregados são WSDL, XML, SOAP etc.

O desenvolvimento de web services requer utilização de uma série de especificações, como, por exemplo, XSD, WSDL, SOAP. O desenvolvedor possui uma variedade de possibilidades para escolhas, tais como, camadas de transporte, políticas de segurança, mecanismos de codificação de mensagens, versão de XML utilizado, entre outros. Para minimizar essas dificuldades e prover o maior nível de interoperabilidade possível, a WS-I[4] – um consórcio composto de um grande número de fornecedores com objetivo de estabelecer as melhores práticas para interoperabilidade entre web services – definiu padrões para implementação de serviços. Seu produto principal é o Basic Profile (BP), que consiste em um conjunto de padrões e boas práticas. O BP auxilia os desenvolvedores a escolherem as configurações de seus web services de forma a garantir interoperabilidade com o maior número de plataformas tecnológicas. Hewitt [2009] sumarizou os padrões e boas práticas menos óbvios presentes no BP. Boa parte da lista indicada por Hewitt foi utilizada no presente trabalho. A seguir, são apresentadas descrições do que foi empregado:

- O elemento <body> de um <SOAP envelope> deve conter exatamente zero ou um elemento filho e deve ser namespace qualified, ou seja, os elementos e/ou tipos utilizados no esquema devem estar de acordo com o namespace especificado como targetNamespace. Não deve conter instruções de processamento ou utilizar DTD (Data Type Definitions);
- Utilize literal message encoding em detrimento de SOAP encoding. Utilizando literal, os elementos e/ou tipos complexos existentes na mensagem são legíveis (literais) tanto para o consumidor quanto para o provedor, estando desacoplada

[4] http://ws-i.org/Profiles/BasicProfile-2.0-2010-11-09.html

de linguagem e codificação. Utilizando encoded, as mensagens são codificadas utilizando SOAP-Encoding, gerando um acoplamento com esta tecnologia;

- Apesar de ser permitido pela especificação SOAP 1.1, não se deve utilizar notação de ponto para redefinir o significado de um elemento <faultcode>. Essa abordagem simplifica o significado do erro. O BP indica que o detalhamento da informação contida no elemento <faultstring> deve ser mantido;
- Apesar de SOAP Action ter a pretensão de auxiliar na indicação da rota de mensagem para uma determinada operação, na prática, os serviços não permitem a utilização de SOAP Action, de forma que toda a informação pertinente é realizada no envelope SOAP e não em cabeçalhos HTTP. Assim, SOAP Action deve ser utilizado somente com uma dica;
- Todas as SOAP Actions devem ser especificadas no WSDL utilizando uma string entre aspas. Se as SOAP Actions não forem especificadas, deve-se utilizar uma string vazia entre aspas;
- Os cookies podem ser utilizados, mas sua utilização não é incentivada porque consumidores que não suportam cookies não estão aptos a consumirem o serviço. Desta forma, apesar do BP permitir sua utilização, essa tecnologia não é utilizada neste trabalho;
- Utilize XML 1.0 porque esta é a única versão que o BP oferece suporte para XML Schema e WSDL;
- Utilize codificação UTF-8 ou UTF-16;
- Tanto a especificação do WSDL quanto o BP recomendam a separação de arquivos WSDL em componentes modulares e criação de arquivos que separam a definição abstrata da definição concreta do WSDL;
- Ao importar um WSDL, deve haver coerência entre os namespaces de ambos os arquivos. Isso significa que o targetNamespace indicado no elemento <definitions> do WSDL importado deve ser o mesmo namespace utilizado no WSDL que o está importando.
- Utilize somente os recursos proporcionados pela especificação do WSDL. A utilização de extensões de WSDL pode implicar na inviabilidade de consumo de serviços por parte de alguns consumidores que não suportam tais extensões;
- O BP não permite o uso de tipos de arrays no WSDL 1.1 devido à diversidade de interpretações de como estes devem ser utilizados. Essa diversidade implica em uma série de problemas de interoperabilidade. Uma solução alternativa simples ao uso de arrays é definir o atributo maxOccurs de um tipo complexo com um valor maior que 0. É possível determinar um número indeterminado utilizando o valor unbounded para maxOccurs;
- Os estilos existentes na construção de um WSDL são Document e RPC. Quando o modelo selecionado é Document, as mensagens de entrada e saída representam documentos, ou seja, referem-se a um parts cujo tipo é element na seção body do WSDL. Quando o modelo selecionado é RPC, as mensagens correspondem aos

parâmetros de entrada e retorno, e deve-se utilizar obrigatoriamente tipos (simples ou complexos). O BP permite a utilização de ambos os estilos;

- Web services são capazes de abstrair o conteúdo de mensagens e operações fora da camada de transporte. Entretanto, apenas SOAP é suportado pelo BP em elementos <binding>.

2.2. Centralização de Esquema

Erl [2008] apresenta o padrão centralização de esquema e defende a definição de um esquema "oficial" para cada conjunto de informações. Contratos de serviços podem compartilhar esses esquemas centralizados. O objetivo é padronizar os tipos de dados em um determinado domínio, seja um processo, uma indústria (financeira, farmacêutica, petrolífera etc.) ou a organização inteira. A centralização do esquema de dados proporciona a eliminação de tipos redundantes. Entretanto, aumenta o acoplamento dos serviços com os tipos de dados comuns centralizados no esquema, o que pode significar um desafio à gestão dos tipos fundamentais.

Esta abordagem garante a interoperabilidade e eficiência dos serviços que participam do processo e diminui a possibilidade de necessidade de redefinição de uma determinada entidade apenas para um ou outro grupo de consumidores. Apesar de não eliminar a necessidade de transformação de dados, essa abordagem a minimiza e possibilita uma governança mais eficiente sobre os dados. Seguindo as melhores práticas de padrão de XML Schema, Hewitt [2009] propõe a utilização de alguns padrões de projeto para esquemas, buscando obter uma mescla de coesão, acoplando, simplicidade e exposição de tipos. São eles: Boneca Russa, Fatia de Salame, Veneziana e Jardim do Éden.

A fim de aproveitar vantagens cruciais da proposta, como uso de múltiplos arquivos, facilidade de reuso entre vários serviços e o baixo acoplamento dos tipos de dados, neste trabalho, foi utilizado o padrão Jardim do Éden a fim de aproveitar o uso de múltiplos arquivos, facilidade de reuso entre vários serviços e o baixo acoplamento dos tipos de dados, porém com adaptações para que possam ser minimizadas as desvantagens enumeradas por Hewitt [2009].

2.3. Contract-first versus code-first

Na abordagem *code-first*, primeiramente é desenvolvida a codificação e, em seguida, gera-se o contrato do serviço com auxílio de um framework. Essa abordagem é mais utilizada devido a sua praticidade [Kalin, 2009]. O desenvolvedor não precisa conhecer e desenvolver a documentação técnica do contrato de serviços (WSDL, XSD e WS-Policy) e, por isso, o desenvolvimento é mais rápido. Contudo, em projetos de médio a grande porte, onde se planeja utilizar SOA ou a mesma já está em execução, essa abordagem pode interferir em princípios básicos da arquitetura como alinhamento com o negócio, interoperabilidade, acoplamento, reuso e durabilidade, como apresentado a seguir. Em *contract-first*, primeiramente elabora-se o contrato do serviço e, em seguida, utiliza-se desse contrato para gerar a codificação correspondente. Como a codificação depende do contrato, observa-se um acoplamento positivo da lógica (codificação) para o contrato (WSDL, XSD e WS-Policy) [ERL, 2009]. A abordagem *contract-first* é menos difundida. Possíveis explicações para esse fato são: maior complexidade de

desenvolvimento devido à necessidade de domínio sobre elementos diferentes da linguagem de programação usualmente utilizada e pouca difusão das técnicas e benefícios da abordagem.

Quando se codifica primeiro, não é possível determinar completamente como o contrato será gerado. Para serviços já disponibilizados, essa abordagem traz uma dependência significativa à tecnologia utilizada para manutenção do serviço. Na prática, significa que se for identificada a necessidade de trocar a tecnologia de desenvolvimento de serviços, será necessária também a alteração do contrato, o que afeta diretamente todos os consumidores destes serviços.

Outra questão é sobre restrições de tipos de dados que podem ser expressas nas estruturas de dados que representam as entidades e algumas dessas restrições exigem desenvolvimento para serem verificadas, ou seja, parte da validação pode ocorrer no XSD e outra parte em uma instância do serviço. O desempenho de SOA torna-se crítico normalmente por causa do tempo de execução dos serviços [Josuttis, 2007]. Assim, qualquer processamento que possa ser evitado deve ser considerado.

3. Provimento de serviços interoperáveis

Para validar os conceitos abordados, foram utilizados os serviços identificados por Azevedo *et al.* [2013]. Em termos de desenvolvimento, cada agrupamento representa um projeto para implementação de um web service, enquanto que cada serviço foi implementado como uma operação do web service.

Os esquemas foram desenvolvidos utilizando o padrão XML Schema – aprovado como especificação pela W3C e incorporado pelo WS-I BP – em sua versão 1.0, utilizando a plataforma Eclipse WTP. O modelo canônico foi empregado neste trabalho considerando um esquema centralizado contendo suas entidades básicas. Além deste esquema central, deve existir um esquema distinto para cada entidade, localizado em um arquivo XSD a parte, representando seus relacionamentos e cardinalidades. Esses esquemas foram chamados de visões. Cada esquema de visão se utiliza do esquema central para manipular os tipos de dados do domínio, maximizando o reuso dos mesmos. A abordagem permite inclusive flexibilidade de que novos esquemas de visão sejam criados, de acordo com a necessidade do serviço. Adicionalmente, as especificações das estruturas das mensagens trocadas entre serviços e seus consumidores possuem apenas informações relevantes à operação, evitando o consumo de recursos desnecessariamente.

Seguindo a abordagem de centralização de esquema proposta, foi desenvolvido um esquema central, chamado ModeloCanonico.xsd, contendo todas as entidades identificadas no modelo canônico. Este esquema conteve apenas as entidades e seus tipos de dados básicos, ou seja, não incluiu relacionamentos. O objetivo é que os serviços possam reaproveitar os tipos de dados e consigam manter a mesma estrutura do modelo canônico, facilitando a manutenção da estrutura. Já o contrato do serviço se utiliza de esquemas que denominamos esquemas de visão, onde são reaproveitados os tipos complexos do esquema central e são mapeados os relacionamentos das entidades utilizadas no esquema de visão. Isto permite que apenas as estruturas necessárias sejam trafegadas, ou seja, mesmo que o modelo de dados conceitual seja todo interligado, somente os dados necessários ao consumo do serviço são trafegados.

O desenvolvimento do WSDL foi realizado manualmente e teve como base as exigências do Basic Profile da WS-I, e o conceito de *contract-first* e suas possibilidades, como modularização e a reutilização dos tipos de dados desenvolvidos. Durante o processo, alguns testes foram realizados, tais como mudanças de nomenclatura e troca de bibliotecas, para verificar que o contrato se mantinha inalterado. Após isso, a ferramenta SOAP UI foi utilizada para realizar a verificação de todas as normas apresentadas pelo WS-I BP e para cada norma de interoperabilidade, o relatório informa o resultado do teste (aprovado, reprovado, aviso, não aplicável, faltando insumo), apresentando uma mensagem detalhada das não conformidades, se estas existirem.

Para a implementação dos serviços, foram utilizadas as especificações presentes na JEE6, edição *enterprise* da linguagem Java, próprias para desenvolvimento de Web Services SOAP. O JAX-WS (*Java API for XML Web Services*) é capaz de incorporar integralmente, através do uso de anotações, todas as capacidades de um WSDL 1.1 a uma interface Java. O uso do JAX-WS se torna fundamental neste trabalho, uma vez que permite utilizar um WSDL já elaborado, tornando possível aplicar a abordagem *contract-first* para desenvolver os serviços, e obter as vantagens observadas anteriormente. Para realizar a ligação (*binding*) entre dados em XML e classes Java [Ort e Metha, 2003] foi utilizado o JAXB (*Java Architecture for XML Binding*) para mapear as classes Java através de anotações, correlacionando tipos e/ou elementos de um XML Schema e atributos de classe Java. Com o domínio do processo bem definido e centralizado, é possível manter as classes JAXB em um único projeto que pode ser utilizado por vários serviços.

Utilizando abordagem *contract-first*, foi possível verificar aderência ao WS-I BP após a confecção do contrato, i.e., os serviços implementados atenderam a todos os critérios do *checklist* do WS-I, dessa forma obtendo um alto nível de interoperabilidade do serviço. Além disso, foram realizados testes envolvendo alterações de operação, entrada de dados e mudanças de tecnologia, e o consumidor do contrato do serviço não precisou ser alterado. Comparado ao uso de *code-first* (sem o uso da nossa abordagem), mesmo mudanças no serviço sem alterar sua interface (i.e., operações e dados de entrada e saída) podem levar a geração de um WSDL e, consequentemente, impactar consumidores do serviço.

A padronização de tipos de dados por domínio (centralização de esquema), desenvolvido a partir do modelo canônico proporcionou a eliminação de tipos de dados redundantes, melhorando a governança sobre os tipos de dados, tornando desnecessária transformações de dados e, consequentemente, melhorando o desempenho e interoperabilidade dos serviços. O desempenho também foi otimizado devido à utilização do padrão Jardim do Éden, associado ao esquema de visões elaborado no projeto. Essas visões proporcionaram diminuição do tráfego de informações em mensagens SOAP, pois apenas os dados definidos no esquema de visão são trafegados.

4. Conclusões

Apesar da abordagem code-first agilizar o desenvolvimento de serviços, esta técnica traz impactos negativos à governança de uma arquitetura orientada a serviços [Ganguly e Goswami, 2011]. Neste trabalho, foi abordada a utilização do padrão de desenvolvimento *contract-first* como alternativa ao *code-first* e as implicações negativas

que a geração automática de contrato traz. Para isso, foi dada sequência ao desenvolvimento dos serviços candidatos identificados e analisados por Azevedo *et al.* [2013], cujos passos são os principais para um ciclo de vida de desenvolvimento de serviços, como o proposto por Gu e Lago [2007].

A implementação realizada da abordagem de desenvolvimento *contract-first* aderente ao WS-I BP permitiu, ao longo do trabalho, avaliar mudanças na lógica de codificação, nomenclatura de métodos e de atributos, e troca de bibliotecas da tecnologia, sem que houvesse impacto nos testes de consumo do serviço, enquanto que simulações com a geração automática do contrato gerassem mudanças impactantes, como por exemplo, no namespace e nome dos atributos. Foi possível perceber também, a possibilidade de existirem diferentes implementações para o mesmo contrato, permitindo atender diferentes características de consumo do serviço. A centralização de esquema, por sua vez, permitiu que as entidades do modelo fossem construídas para possuir maior semântica, incorporando as restrições que o domínio expunha na modelagem, tais como valores padrão, mínimos e máximos. Isto também traz benefícios ao consumidor do serviço, possibilitando validar a mensagem que será enviada ao serviço antes da mensagem chegar ao seu destino.

Com este trabalho, conseguimos demonstrar as possibilidades de uso da abordagem *contract-first* comparado ao uso de *code-first*, utilizando tecnologias modernas e utilizadas em larga escala na indústria, apresentando os benefícios ao em utilizar *contract-first.*

5. Referências

AZEVEDO, L. G., SANTORO, F. M., BAIÃO, F., DIIRR, T., SOUZA, A., SOUZA, J. F., and SOUSA, H. P. (2013) "A method for bridging the gap between business process models and services". iSys: Revista Brasileira de Sistemas de Informação, 6(1):62–98.

ERL, T. (2005) Service-Oriented Architecture: concepts, technology, and design. Prentice Hall Englewood Cliffs.

ERL, T. (2008) SOA: Principles of Service Design, Prentice Hall Upper Saddle River.

GANGULY, K., GOSWAMI, P. (2011) "Developing web services, Part 1: Developing the code and contract first approach web service with Axis2". IBM developerWorks. http://www.ibm.com/developerworks/library/ws-devaxis2part1/, acesso 27/10/2014.

GU, Q., LAGO, P., (2007) "A stakeholder-driven service life cycle model for SOA". In: Proc. of 2nd International Workshop on Service Oriented Software Engineering.

HEWITT, E. (2009) Java SOA Cookbook. O'Reilly Media inc.

JOSUTTIS, N. M. (2007) SOA in Practice: The Art of Distributed System Design, O'Reilly Media, Inc.

KALIN, M. (2009) Java Web Services: Up and Running, 2009. O'Reilly Media, Inc.

ORT, E. METHA, B. (2003) "Java Architecture for XML Binding (JAXB)". Sun Developer Network. <http://www.oracle.com/technetwork/articles/javase/index-140168.html>, acesso 03/10/2014.

O Uso de *WebSockets* no Desenvolvimento de Sistemas Baseados em uma Arquitetura *Front-end* com API

Guilherme S. A. Gonçalves, Paulo Henrique O. Bastos, Daniel de Oliveira

Instituto de Computação – Sistemas de Informação – Universidade Federal Fluminense (UFF) – Caixa Postal 24.210-240 – Niterói – Rio de Janeiro – Brasil

galves@id.uff.br, ph_ouverney@id.uff.br,danielcmo@ic.uff.br

Resumo. *Este artigo apresenta um relato de experiências sobre o uso de WebSockets como uma tecnologia capaz de atender um novo tipo de tendência gerada pelas demandas dos sistemas de informação que necessitam respostas em tempo real. Também é discutido nesse artigo como o WebSocket pode ser integrado a uma arquitetura RESTful, criando um meio de comunicação em módulos que são utilizados para prover serviços aos usuários. Assim, nesse artigo, discutimos as vantagens e as desvantagens do uso dessa tecnologia em relação ao padrão de arquitetura atual para o desenvolvimento de aplicações Web.*

1. Introdução

Em muitos dos sistemas de informação atuais a pronta resposta a ações do usuário tem se tornado um requisito quase que obrigatório. Dependendo do tipo de aplicação, não é recomendado que o usuário espere muito tempo até que se tenha um *feedback*. Um exemplo de aplicação com essa necessidade é o TaxiVis (Ferreira *et al.*, 2013), uma aplicação desenvolvida pela NYU (Universidade de Nova York) onde agentes do governo são capazes de mapear dados de movimento de taxis na cidade de Nova Iorque em tempo real. Para que se tomem ações rápidas, não faz sentido que a resposta do sistema seja demorada. Esse problema torna-se ainda mais complexo em aplicações *Web,* dado o tipo de arquitetura das aplicações. Assim, é indispensável nos dias de hoje que o atraso de comunicação entre o servidor e o cliente seja quase nulo e o limite de atraso sempre seja reduzido (Horey e Lagesse, 2011). Exemplos práticos são aplicações de conversas *online* ou um sistema de compra de ações, onde nesses casos queremos que as informações sejam disponibilizadas para nós o mais rápido possível.

Tradicionalmente, as aplicações *Web* tem sido desenvolvidas utilizando arquitetura MVC (*Model-View-Controller*), que é um padrão de projetos amplamente utilizado nas práticas de programação. Este é composto por três módulos: o modelo, onde são implantadas as regras de modelagem e de negócio; a visão, onde estão localizadas as interfaces no qual as interações com os usuários são realizadas; e por último o controlador, no qual são geridas as comunicações e conexões entre o modelo e a visão. (Armeli-Battana, 2012).

O funcionamento desse padrão de programação ocorre de modo que uma requisição é realizada por meio da visão ao controlador que executa um processo, podendo invocar um ou mais serviços para entregar à visão os dados requisitados ou executar uma determinada ação. O controlador é responsável por gerenciar o fluxo das páginas e persistir em sessão os objetos que serão acessados. Uma das grandes vantagens do modelo MVC é que ele é de fácil implementação e aceitação no mercado,

já que é o padrão *de facto* na indústria. Por seu claro entendimento, por prover uma organização do código e facilitar a programação, muitos entusiastas do padrão disponibilizam materiais, explicações, tutoriais e cursos sobre o tema, que acaba difundindo ainda mais o padrão. Além de ter uma compreensão transparente do fluxo, por ele estar bem explícito.

Apesar das vantagens anteriormente apresentadas, o MVC tem algumas desvantagens na sua implementação. As regras de negócio se encontram fixas no código e não podem ser compartilhadas para outras aplicações, como por exemplo aplicações móveis em celulares e *tablets*. Isto torna a aplicação "engessada" para um determinado tipo de arquitetura, inviabilizando reutilização de código e dos serviços. Para resolver esse problema, seria necessário implementar uma API para que esses serviços estivessem disponíveis para outras arquiteturas. Isso faz com que o trabalho seja dobrado. Além disso, a própria definição do MVC já faz com que a troca de mensagens insira um retardo de comunicação que em alguns tipos de aplicação pode não ser aceitável.

Essas limitações do MVC vem sendo suplantadas nos últimos anos graças às novas tecnologias que surgem para prover novos tipos de serviços para áreas que demandam cada vez mais processamento e transferência de dados com atrasos mínimos e garantias de segurança. Um exemplo é a arquitetura *Front-end* que consome serviços implementados por uma API. Nessa arquitetura, o *Front-end* pode utilizar o modelo MVVM - *Model-View-ViewModel*, que pode ser visto na Figura 1, implementado, por exemplo, pelo AngularJS. A API que vai prover os serviços pode ser desenvolvida em qualquer linguagem ou até em múltiplas linguagens, sendo o mais importante a forma como será realizada a comunicação entre elas e os módulos interligados. Elas podem utilizar o formato REST ao invés do SOAP para a comunicação, utilizando o JSON (Richardson e Ruby, 2008) no formato de mensagens trocadas. Na Figura 2 é exibido um exemplo de mensagem no formato JSON que pode ser trocada pelas aplicações.

Assim como o MVC, a arquitetura *Front-end* com API possui vantagens e desvantagens associadas. A maior vantagem dessa arquitetura *Front-end* com API é a separação das obrigações, onde o desenvolvedor define que as regras de negócio ficam na API, o que possibilita seu compartilhamento por diversas aplicações e viabiliza as práticas de reutilização de código, evitando o retrabalho.

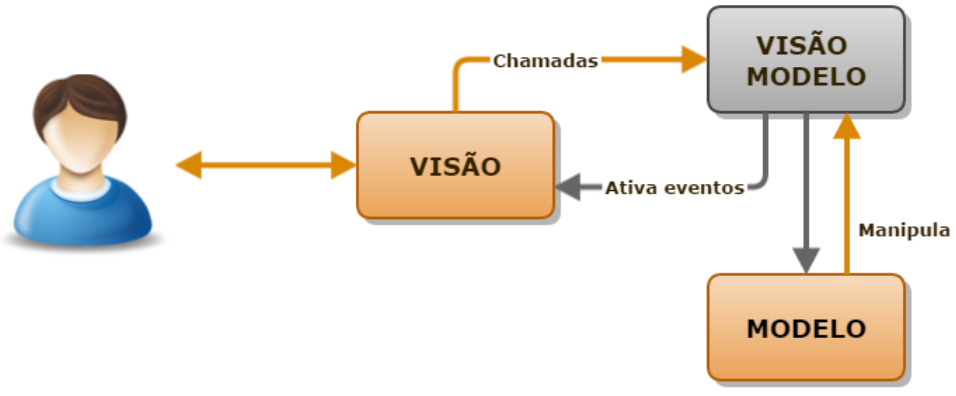


Figura 1. Um esquema conceitual da arquitetura Model-View-ViewModel.

Já a maior desvantagem dessa arquitetura é o fato de ser ainda pouco conhecida e adotada, pois há pouca divulgação e oferta de aprendizado. Outra desvantagem conhecida está relacionada ao esforço de se implantar um serviço de segurança, onde o tráfego de mensagens JSON deve ser criptografado antes de ser enviado, pois há o risco das mensagens serem interceptadas durante a comunicação. Esse requisito de segurança é um ponto crucial na aplicação dessa arquitetura. Além disso, ainda existe o problema de retardo de comunicação que deve ser tratado para que possa ser utilizado para desenvolver aplicações que demandem respostas em tempo real.

Esse artigo apresenta algumas discussões e soluções para a implantação e o uso da arquitetura *Front-end* com API por meio do uso de *WebSockets* de forma que ela possa ser mais difundida como uma alternativa ao padrão MVC.

```
{
    id:2,
    nome: Daniel de Oliveira,
    endereco: {
                    id: 2,
                    estado: Rio de Janeiro,
                    cidade: Niterói,
                    rua: Rua Passo da Pátria,
                    numero: 156
                }
}
```

Figura 2. Exemplo de uma mensagem no formato JSON.

2. Princípios da Arquitetura *Front-end* com API com uso de *WebSockets*

Para que possamos suplantar os problemas apresentados anteriormente da arquitetura *Front-end* com API, temos que adaptar alguns pontos no seu uso. Em relação aos atrasos de comunicação e de forma a obter latências de conexões quase nulas ou em tempo real entre clientes e servidores é necessário que utilizemos mecanismos os quais oferecem mais recursos do que o serviço oferecido pelo protocolo HTTP.

Os *WebSockets* surgiram para prover esse tipo de serviço com retardo reduzido, o qual o HTTP não oferece apoio atualmente. O *WebSocket* é um padrão ainda em desenvolvimento que está sob os cuidados da IETF (*Internet Enginnering Task Force*) e uma API padrão para implementação do protocolo está em processo de formalização pela W3C (*World Wide Web Consortium*) para que os navegadores ofereçam apoio ao protocolo. O suporte dos navegadores ao protocolo já se encontra em desenvolvimento e disponibilização ao usuário final, entretanto utilizar as versões mais recentes dos navegadores será indispensável para quem quiser desenvolver sistemas de informação utilizando essa arquitetura e usufruir do serviço que está em constante aprimoramento. (Cutting Edge, 2014).

Por implementar um canal de transmissão bidirecional e por ter um cabeçalho menor, o *WebSocket* supera o HTTP em relação ao número de requisições feitas e no tamanho das mensagens. Podemos ver na Figura 3, que o *handshake* do *WebSocket* é mais simples, pois é feito apenas um por conexão, enquanto no HTTP é realizado por

requisição. Após a conexão ser estabelecida, somente são trocadas mensagens e sem necessidades de um cabeçalho extenso.

Em relação ao requisito de segurança, podemos utilizar o *WebSocket secure* (WSS) como uma alternativa de meio de transmissão seguro em relação ao *WebSocket* convencional (Richardson e Ruby, 2008).

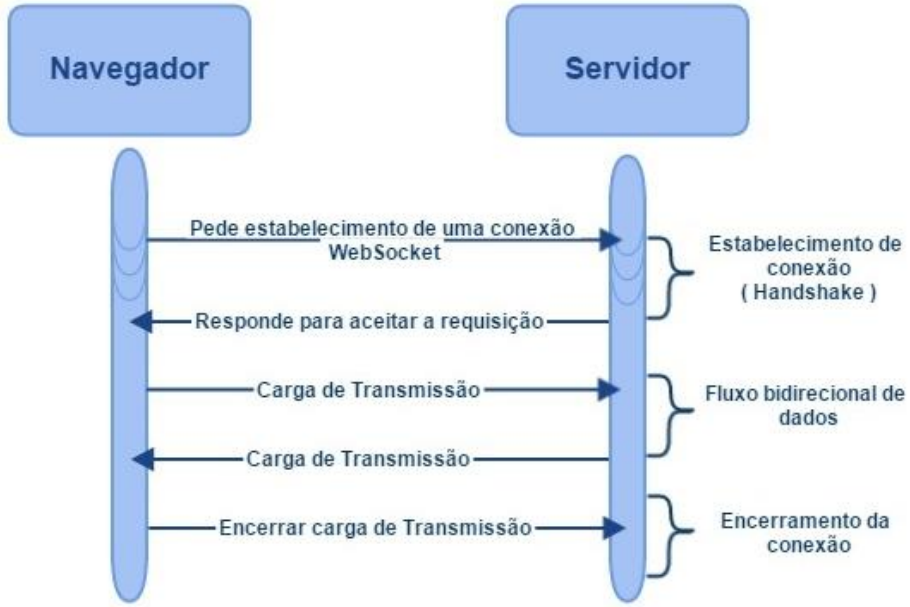


Figura 3 Handshake no WebSocket.

A Figura 4 e a Tabela 1 apresentam resultados comparativos entre um sistema de informação *Web* utilizando o REST por HTTP e a outra utilizando *WebSocket*. Os sistemas que utilizam *WebSocket* são capazes de processar mais mensagens em uma determinada faixa de tempo em comparação aos sistemas que são baseados no protocolo HTTP, dessa forma aumentando a vazão de informação que o sistema é capaz de processar.

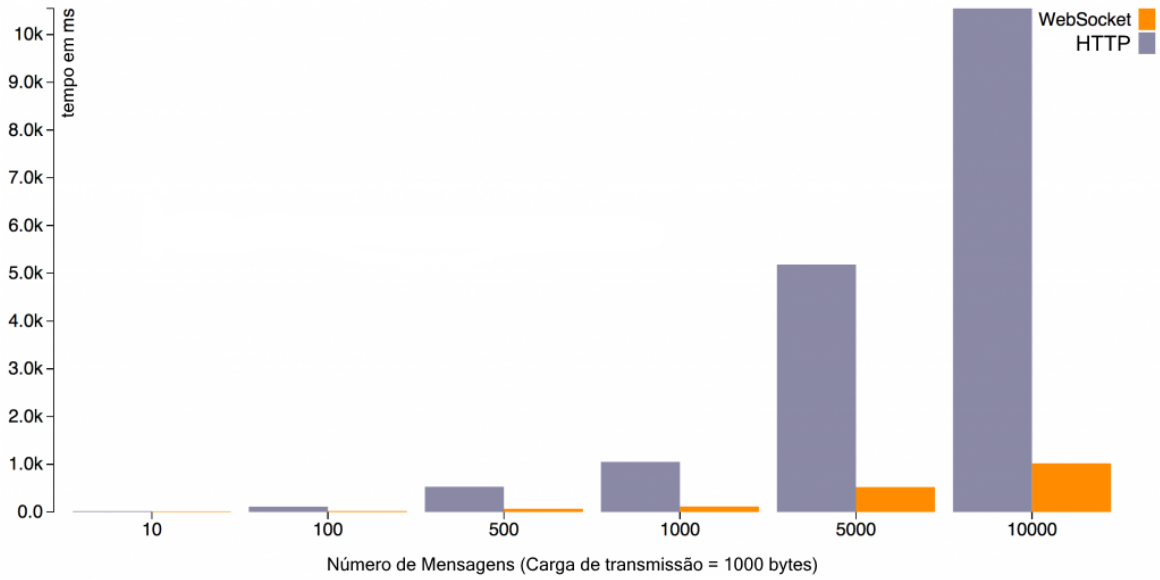


Figura 4. Gráfico comparativo entre o *WebSocket* e a arquitetura REST utilizando o protocolo HTTP em relação ao números de mensagens enviadas adaptado de (Planet JBoss, 2014).

Tabela 1. Tabela comparativa entre o *WebSocket* e a arquitetura REST utilizando o protocolo HTTP em relação ao números de mensagens enviadas (Planet JBoss, 2014).

Mensagens	HTTP (em ms)	*WebSocket* (em ms)	Proporção
10	17	13	1,31
100	112	20	5,60
500	529	68	7,78
1000	1.050	115	9,13
5000	5.183	522	9,93
10000	10.547	1.019	10,35

Outro problema do uso de *WebSockets* é que o seu cabeçalho não é extenso e somente disponibiliza os métodos "*OnOpen*", "*OnMessage*" e "*OnClose*", não é possível identificar que tipo de requisição é realizada através de uma URI, como é feito em aplicações *RESTful* com HTTP (Fette e Melnikov, 2014). Devido a essa limitação, a solução proposta nesse artigo é que o tipo de requisição realizada esteja na mensagem e não no cabeçalho do protocolo. Assim, na Figura 5 é exemplificada uma mensagem JSON sendo enviada para uma conexão já previamente aberta. Nota-se que o campo "*method*" é responsável por informar à API que tipo de requisição é realizada, podendo aceitar os valores: "*get*", "*create*", "*update*", "delete" e "*findAll*".

```
{
        "callback_id": "1",
        "method": "get",
        "data": {
                "type": "user",
                "id": "9"
                }
}
```

Figura 5. Exemplo de uma mensagem no formato JSON utilizando a metodologia para criar uma arquitetura RESTful.

É importante abrir uma conexão por módulo/domínio e cada conexão aberta deve ser responsável por receber os tipos de requisições para o seu módulo/domínio. Na Figura 6, é apresentada a arquitetura da solução proposta, onde os módulos/domínios

são usuário e empresa, e os métodos abaixo do módulo/domínio serão acessados por meio de sua conexão responsável. A conexão somente irá se fechar quando não for mais necessário acessar qualquer serviço de usuário ou empresa.

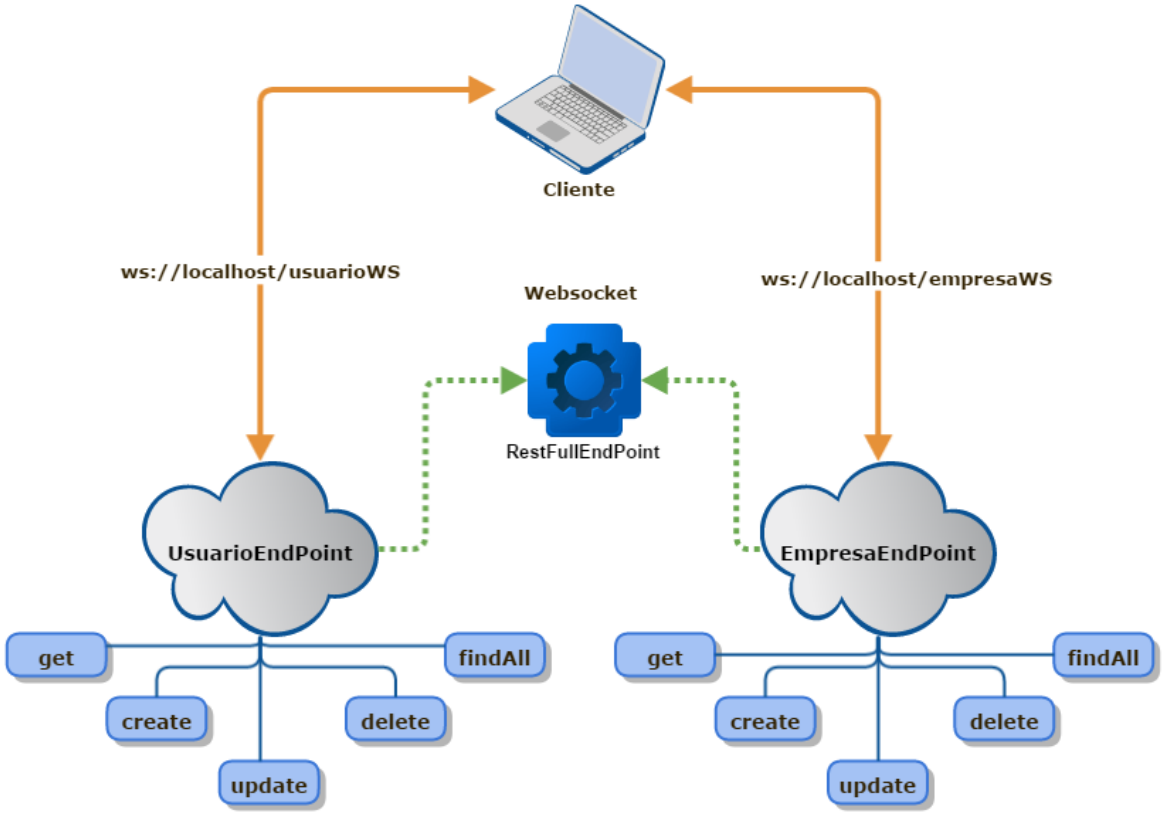


Figura 6. Exemplo de conexão do cliente com dois módulos utilizando uma comunicação *WebSocket*.

3. Conclusões

Esse artigo tem como objetivo apresentar uma nova opção de arquitetura de sistemas de informação em relação ao tradicional modelo MVC. Apesar de ainda possuir uma série de limitações e problemas envolvidos, essa arquitetura se mostra promissora para o desenvolvimento de sistemas de informação que necessitem ser multiplataforma e que necessitem de respostas em tempo real. A utilização de *WebSockets* se mostra uma solução promissora para suplantar os problemas da arquitetura *Front-end* com API e certamente deve ser alvo de investigação tanto por parte da academia quando por parte da indústria nos próximos anos.

4. Referências Bibliográficas

Nivan Ferreira, Jorge Poco, Huy T. Vo, Juliana Freire, Cláudio T. Silva. (2013): Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips. IEEE Trans. Vis. Comput. Graph. 19(12): 2149-2158

James Horey, Brent Lagesse. (2011) Latency Minimizing Tasking for Information Processing Systems. ICDM Workshops., p. 167-173

Sebastiano Armeli-Battana. (2012), MVC Essentials [Kindle Edition], developer press.

Padrões de Projeto - O modelo MVC - Model View Controller ([S.d.]). http://www.macoratti.net/vbn_mvc.htm, [acessado em Setembro de 2014].

.NET - Apresentando o padrão Model-View-ViewModel ([S.d.]). http://www.macoratti.net/11/06/pp_mvvm1.htm, [acessado em Setembro de 2014].

Richardson, L. e Ruby, S. (2008), *RESTful Web Services*. O'Reilly Media, Inc.

Kaazing Developer Network ([S.d.]). http://developer.kaazing.com/documentation/jms/4.0/security/c_sec_https_wss.html, [acessado em Setembro de 2014].

Planet JBoss, REST vs WebSocket Comparison and Benchmarks | Planet JBoss Developer ([S.d.]). https://planet.jboss.org/post/rest_vs_websocket_comparison_and_benchmarks, [acessado em Setembro de 2014].

Fette, I. e Melnikov, A. ([S.d.]). The WebSocket Protocol. http://tools.ietf.org/html/rfc6455, [acessado em Setembro de 2014].

Castillo, I. e Pascual, V. ([S.d.]). The WebSocket Protocol as a Transport for the Session Initiation Protocol (SIP). http://tools.ietf.org/html/rfc7118, [acessado em Setembro de 2014].

Cutting Edge - Compreendendo o poder dos WebSockets ([S.d.]). http://msdn.microsoft.com/pt-br/magazine/hh975342.aspx, [acessado em Setembro de 2014].

De Col, A. and Nesello, F. A. (2014). Aplicativo web com JSF 2.0 e PrimeFaces para gerenciamento de requisitos de software.

A adoção de uma arquitetura de sistemas orientada a serviços para integração de sistemas de informação em um instituição pública de ensino

Bruno L. Sousa[1], **Angelo B. N. Júnior**[1], **Daves M. S. Martins** [1]

[1]Departamento Acadêmico de Ciência da Computação – Instituto Federal de Educação, Ciência e Tecnologia Sudeste de Minas Gerais – Câmpus Rio Pomba

{bruno.luan.sousa, nevessangelo}@gmail.com, daves.martins@ifsudestemg.edu.br

Abstract. *The web, in actuality, is a platform used by many applications. An application to problems in its structure can not provide accurate information and makes it difficult to communication and information sharing between systems. At the Federal Institute of Education, Science and Technology Southeast of Minas Gerais - Campus Rio Pomba there are several systems that do not communicate. The information is decentralized, which cause the "Islands of Information". Therefore, the purpose of this article is to show the use of Service Oriented Architecture (SOA) technology to integrate existing systems within the institution. Therefore, systems developed in different programming languages pro- can communicate providing sharing and reuse of information.*

Resumo. *A web, na atualidade, é uma plataforma utilizada por muitas aplicações. Uma aplicação com problemas em sua estrutura não consegue fornecer informações precisas e torna difícil a comunicação e o compartilhamento de informações entre sistemas. No Instituto Federal de Educação, Ciência e Tecnologia Sudeste de Minas Gerais - Câmpus Rio Pomba existem vários sistemas que não se comunicam. As informações ficam descentralizadas, o que causam as "Ilhas de Informações". Portanto, a finalidade deste artigo é mostrar o uso da tecnologia Arquitetura Orientada a Serviços (SOA) para integrar os sistemas existentes na instituição. Assim, sistemas desenvolvidos em linguagens de programação diferentes poderão se comunicar proporcionando um compartilhamento e reaproveitamento das informações.*

1. Introdução

Com o desenvolvimento da tecnologia o principal aspecto nos dias atuais é a informação. Ela possibilita a comunicação entre pessoas e o estabelecimento de novas relações. Sua utilização é essencial para os sistemas computacionais, pois são a base de um bom funcionamento. Quando armazenadas e processadas de forma correta proporcionam agilidade e eficiência em diversas atividades com um ganho imenso de tempo em um mundo onde o mesmo é escasso.

No meio acadêmico não é diferente. Na atual realidade das Universidades muitas não conseguem processar esses conjuntos de informações de forma unificada. A falta de uma integração permite as informações ficarem espalhadas e serem alocadas em várias bases de dados. Segundo [LI et al. 2010], quase todas as Universidades não possuem sistemas integrados e geram assim as chamadas “Ilhas de Informações”.

Ainda, essa falta de integração gera uma redundância das informações, e passa a ser um grande problema, pois a comunicação não ocorre de forma clara e objetiva. As mesmas são cadastradas várias vezes, onde para cada sistema é necessário efetuar um cadastro, que poderão sofrer alterações e ficarem muitas das vezes inutilizáveis ou inconsistentes.

2. Problemas

O cenário apresentado na Seção 1 também está presente no Instituto Federal de Educação, Ciência e Tecnologia Sudeste de Minas Gerais – Câmpus Rio Pomba (IFSudesteMG - Câmpus Rio Pomba). Existe uma grande quantidade de sistemas que atendem a diversos setores. Cada setor tem a liberdade projetar e controlar esses desenvolvimentos sem qualquer padronização tornando-os isolados dos demais existentes. Cada sistema possui o seu próprio banco de dados. Assim, as informações dos usuários ficam espalhadas e se tornam redundantes, uma vez que as informações não são compartilhadas. Esse mesmo cenário ocorre em [ZHOU 2011] e [MIRCEA 2012].

No câmpus Rio Pomba, também não há uma interface unificada. Em [NETO et al. 2014], algumas medidas que visam resolver esse problema foram propostas, porém limita os desenvolvimentos à uma única linguagem, o PHP.

Atualmente, esses sistemas estão desenvolvidos em uma arquitetura estruturada, mostrada na **FIGURA 1**. Possui uma comunicação cliente-servidor, onde suas informações são requeridas pelo cliente e são acessadas ou processadas direto no banco de dados de acordo com a regra de negócio. Esta arquitetura está descentralizada, contudo, não há separações e reaproveitamento de funções. Para cada função desenvolvida no site necessita-se escrevê-la totalmente. Assim, os sistemas ficam isolados e as informações ficam restritas.

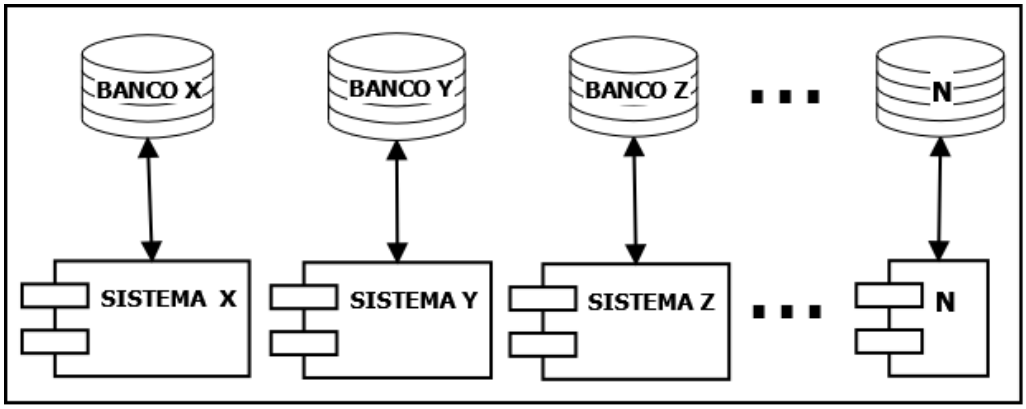


Figura 1. Arquitetura atual dos sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais – Câmpus Rio Pomba

3. Solução Apresentada

Existem várias tecnologias que podem ser usadas para amenizar os problemas discutidos na Seção 2. Dentre elas encontra-se o Service Oriented Architecture (SOA), um novo paradigma que despontou na atualidade para o desenvolvimento de sistemas. Segundo [CUNHA et al. 2014] o SOA é uma arquitetura que refere-se a um planejamento

de estratégias que permitem as traduções e compartilhamento de algumas funcionalidades das aplicações e assim, mantêm nas inter-relacionadas. Além disso, os serviços webs (web services), juntamente com o SOA, vieram com o intuito de integrar aplicações web, onde são disponibilizados XMLs que permite acesso de várias aplicações [GAO and WANG 2013]. Há uma grande melhora na comunicação entre várias linguagens, uma vez que todas conseguem entendê-lo. Assim, os sistemas conseguem compartilhar informações e dão fim às ilhas de Informações e redundâncias existentes, pois, informações já cadastradas são disponibilizadas para várias aplicações, como em [ZHOU 2011].

Ainda, existe outra tecnologia que é muito usada nos dias atuais como forma de autenticação. Segundo [AUTHENTICATIONWORLD.COM 2014], o Single Sing On é uma tecnologia que permite utilizar o mesmo id (login) e senha para várias aplicações. Dessa maneira, pode-se fazer um controle múltiplos de acessos do usuário com apenas um único login e não deixar o mesmo à mercê de vários senhas e logins heterogêneos, que muitas das vezes geram o esquecimento de algumas delas. Pode-se ver a utilização do Sigle Sign On em trabalhos como [WANG and ZHANG 2014], [GAO and WANG 2013] e [WANG et al. 2012].

Assim, a tecnologia SOA foi adotada para a comunicação e compartilhamento entre aplicações e o SSO (Single Sign On) para a autenticação. Uma solução que se baseia na tecnologia de disponibilização de serviços será a responsável por armazená e prover as informações aos demais sistemas, independente da linguagem de desenvolvimento. Outro aspecto relevante a ser levantado diz respeito à segurança da informação, onde esta, estará protegida por uma camada de serviços e sem redundâncias. Além da disponibilização de serviços haverá uma infraestrutura de controle de usuários e suas permissões.

Basicamente a nova arquitetura se baseará em três partes, como mostra a **FIGURA 2.**.

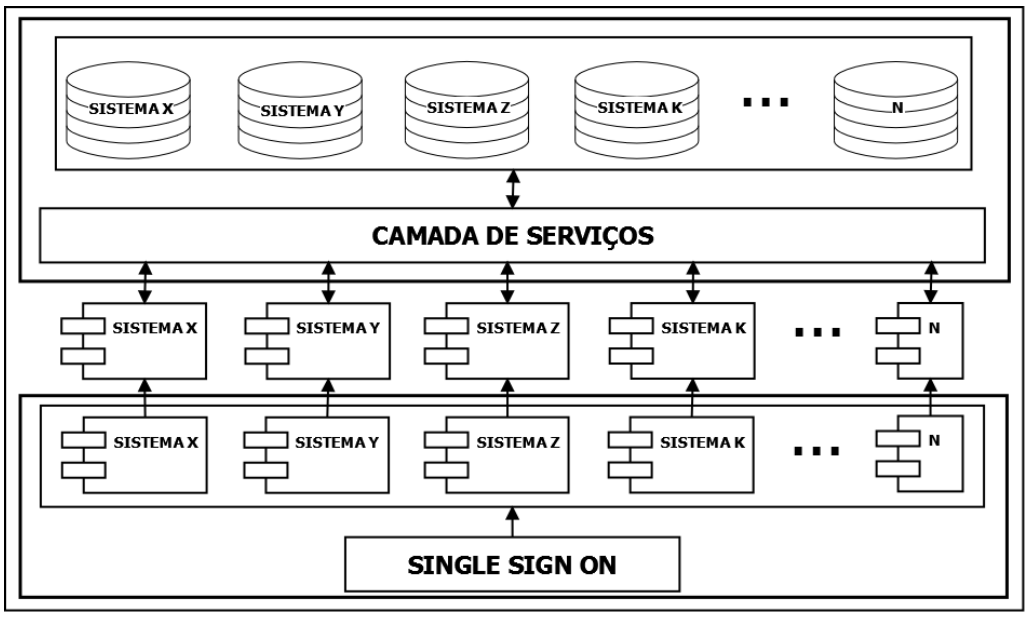


Figura 2. Arquitetura proposta para o Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais – Câmpus Rio Pomba

A primeira parte será a plataforma principal, na qual terá autenticação unificada (Single Sign On), que quando utilizada pelo usuário, o mesmo será levado para uma página com as aplicações, de acordo com suas permissões. Será uma espécie de portal onde terá as aplicações disponíveis para acesso.

A segunda parte é a própria aplicação. Ao clicar em uma das opções disponibilizadas no nosso portal o usuário será redirecionado direto para a aplicação ou sistema escolhido, onde entra como usuário já logado de acordo com suas permissões.

A terceira parte são as camadas de serviços e os bancos de dados. Como já foi mencionado, os web services são os responsáveis por disponibilizar todos os serviços. Com isso, a plataforma conseguirá fazer a comunicação entre várias linguagens diferentes através do XML gerado. Assim, a plataforma não fica restrita a apenas sistemas desenvolvidos em um tipo de linguagem. Além disso, esses serviços são responsáveis por manipular as informações no banco de dados, onde passa-se pela camada de serviço antes de acessá-lo.

4. Desenvolvimento e Implementação da Proposta

Após esse estudo inicial começou-se a implantação dessa nova arquitetura, na qual dividiu-se em duas partes. A primeira visava a construção da plataforma de serviço e a segunda à integração do sistema de Educação Física, que foi o primeiro do IFSudesteMG - Câmpus Rio Pomba integrado nessa nova arquitetura.

4.1. Desenvolvimento do Portal e da plataforma de serviços

A segurança nos níveis de acesso são pontos cruciais para um bom funcionamento. A partir dela consegue-se impedir que muitas pessoas acessem conteúdo que são sigilosos para seu referido nível de acesso. Segundo [LOPES and MARCHI], a ferramenta Spring Security da Framework Spring, mantém uma estrutura de controle, e consiste de uma autenticação forte e altamente personalizável, a qual garante uma boa proteção para seu sistema ou aplicativo. Em [MEZAROBA et al. 2014], pode-se ver a utilização dessa ferramenta, onde os autores utilizam-na para a validação e autenticação em um sistema de Registro Eletrônico de Pacientes. Baseado nesses trabalhos, resolveu- se adotar a ferramenta Spring Security da Framework Spring para proteção na hora da autenticação.

Como o objetivo é desenvolver uma plataforma unificada, obrigatoriamente será usada a tecnologia de SSO para uma autenticação unificada. Nessa autenticação unificada fez-se o uso do token, um identificador do usuário que conterá todas as informações do mesmo em qualquer sistema. Através dele consegue-se saber que permissão o usuário possui e quais sistemas ele terá acesso. Para que a autenticação fique centralizada nessa plataforma é gerado um cookie com um token. Mas como uma medida de segurança, cada vez que o usuário logar é gerado um token diferente para que o sistema não seja burlado. Os serviços desenvolvidos nessa plataforma foram: serviços de usuários, serviços de logon, serviços de logoff e recuperação de dados.

- Nos serviços de usuários foram feitos vários serviços de acordo com a função e o tipo de usuário a ser executado. A função vai dizer se será um cadastro, ou uma alteração do seu perfil. O tipo de usuário vai dizer qual nível de acesso o usuário possui, uma vez que possuem restrições e liberação de acesso de acordo

com o tipo do mesmo. Os níveis irão variar de acordo com os sistemas, porém haverá um que será comum para todos: o administrador. Esse usuário já será previamente cadastrado e terá o nível máximo de acesso que qualquer outro tipo de usuário não poderá alcançar.

- Nos serviços de logon utilizou-se a ferramenta Spring Security da Framework Spring, supracitada no início desta subseção, na qual utiliza um arquivo de configuração, que quando feita uma requisição de logon, esse arquivo é acionado e repassa a ação para o serviço de logon responsável por conferir se o usuário possui algum acesso nessa plataforma. Caso possua, é gerado um token que através de um outro serviço é gravado na base de dados. Esse token será utilizado para identificar o acesso do usuário a qualquer aplicação integrada nessa plataforma.

- Nos serviços de logoff, o arquivo de configuração da ferramenta Spring Security da Framework Spring redimensiona para o serviço, no qual será passado o token do usuário e o mesmo se encarregará de excluir o token na base de dados e destruir o cookie do navegador encerrando o acesso do usuário na plataforma.

- Nos serviços de recuperação de dados são feitas recuperação dos dados, onde é passado como parâmetro o token e retornado todos os dados do usuário que está logado no site. Esses serviços serão muito utilizados quando o usuário for alterar seu perfil, onde a plataforma terá que saber todos os dados do mesmo.

Assim, através desses serviços foi construído o módulo central, o qual será a parte fundamental para toda a integração. Nela haverá os links para todos os sistemas que ao ser clicado redireciona o usuário para sua escolha, onde já estará autenticado com sua devida permissão de acesso.

4.2. Implantação do Sistema de Educação Física na nova arquitetura

Esse sistema tem por objetivo o gerenciamento de avaliação e prescrição de atividades físicas no câmpus, assim, minimizar todo o trabalho de gerenciamento e armazenamento das atividades físicas e medidas corporais dos alunos. Para que sua a integração fosse realizada, foi necessário reescrevê-lo e reorganizá-lo dentro dessa nova plataforma, onde foram desenvolvidos serviços de busca de informação dos alunos, serviços para armazenar as atividades dos alunos também foram necessários, serviços de cadastramento e busca de avaliações físicas, serviços de gerenciamento de questionário, dentre outros, que estabeleceram a comunicação deste com o módulo central desenvolvido.

Esse sistema e o módulo central de autenticação dessa nova arquitetura estão em fase de implantação, em uso com alguns usuários selecionados para testes do sistemas, onde serão observados a comunicação e o compartilhamento das duas aplicações.

5. Conclusão

Tendo em vista o objetivo de unificar os dados e sistemas do IFSudesteMG - Câmpus Rio Pomba, foi feito um estudo da arquitetura SOA juntamente com a tecnologia de web service, SSO e a ferramenta Spring Security da Framework Spring para que pudesse ser construído o módulo central dessa nova arquitetura.

Ainda, foi implantado o Sistema de Educação Física dentro dessa nova arquitetura SOA, onde pode-se perceber que todas as expectativas criadas em cima da mesma foram atendidas e grandes resultados foram atendidos, dentre os quais encontra-se o compartilhamento de informações em base de dados distintas, a autenticação única e comunicação entre as aplicações de diferentes linguagens.

Para integrações e trabalhos futuros, serão feitas análises nos sistemas a serem integrados com o objetivo de localizar todos os acessos a seu banco de dados, que atualmente possuem um acesso direto, e reescrevê-los em formas de serviços para que este possa comunicar-se com todos os outros já existentes dentro dessa nova arquitetura.

Portanto, a adoção de uma nova arquitetura orientada a serviços para a integração dos sistemas no Câmpus Rio Pomba trouxe grandes benefícios, uma vez que as informações ficaram mais organizadas e sem redundâncias, além de uma única autenticação para estes acessos dentro do câmpus.

Referências

AUTHENTICATIONWORLD.COM (2014). Single sign on. `http://www.authenticationworld.com/Single-Sign-On-Authentication/`. [Acessado em Agosto de 2014].

CUNHA, M. X., JÚNIOR, M. F. S., and DORNELAS, J. S. (2014). O uso da arquitetura soa como estratégia de integração de sistemas de informação em uma instituição pública de ensino. *SEGeT – Simpósio de Excelência em Gestão e Tecnologia*.

GAO, X. and WANG, L. (2013). Research of digital campus architecture based on soa. *The 8th International Conference on Computer Science Education (ICCSE) (April 26-28). Colombo, Sri Lanka, v. MoC2.4, p.1320-1323.*

LI, M., WANG, L., HAM, X., and JIANG, J. (2010). Research on digital campus of higher colleges and its management platform. *International Conference on Educational and Information Technology (ICEIT), v. 3, p. 471-475*.

LOPES, G. B. and MARCHI, K. R. C. Segurança no desenvolvimento web utilizando framework spring security. `http://web.unipar.br/~seinpar/2013/artigos/Guilherme%20Baiestero%20Lopes.pdf`. [Acessado em Agosto de 2014].

MEZAROBA, W. G., MENEGON, M. P., and NICOLEIT, E. R. (2014). "registro eletrônico de paciente em uma uti: Comunicação, interação com dispositivos móveis e previsão de expansibilidade". *Congresso Brasileiro De Informática Na Saúde, XI. Anais.*

MIRCEA, M. (2012). Soa adoption in higher education: a practical guide to service-oriented virtual learning environment. *World Conference on Learning, Teaching Administration (WCLTA). Science Direxť, v. 31, p. 218-223.*

NETO, J. M., DIAS, R. A., REIS, G. H. R., and COELHO, F. M. (2014). Framework for development of cms for systemic directors. *11TH CONTECSI – Internacional Conference on Information Systems and Technology Management, University of Sao Paulo, p. 1-28.*

WANG, G., XU, G., and ZHANG, Z. (2012). Research on data synchronization and integration platform based on soa. *Journal of Convergence Information Technology(JCIT) Volume 7, Number 13, July 2012*.

WANG, G. and ZHANG, Z. (2014). "design of soa-based university financial information portal". *Soft science research program of Henan province*.

ZHOU, L. (2011). Study on scheme for university information system integration based on soa. *Economics Management College, Nanchang Hangkong University Nanchang. Jiangxi, China, 4p.*

Seleção de Parâmatros para Classificação de Faltas do Tipo Curto-Circuito em Linhas de Transmissão

Jefferson Morais[1], Yomara Pires[2], Fernando dos Santos[2], Waldemar Neto[2], Aldebaro Klautau[1]

[1]Programa de Pós-Graduação em Ciência da Computação
Universidade Federal do Pará (UFPA)
Caixa Postal 479 – 66.075-110 – Belém – PA – Brasil

[2]Faculdade de Computação– Universidade Federal do Pará (UFPA)
Caixa Postal 479 – 68.746-360 – Castanhal – PA – Brasil

{jmorais, yomara, aldebaro}@ufpa.br, fernandojj2@gmail.com, netolandy@hotmail.com

Abstract. *This paper proposes a methodology to feature selection for fault classification in transmission lines. Since most work on the classification of events in power systems is concerned with an event classified according to morphology of the corresponding waveform. An important and yet most difficult problem is the classification of the cause of the event. Thus, this paper uses different techniques for pre-processing (front-ends), feature selection algorithms and machine learning techniques to classification of short-circuit faults.*

Resumo. *Este trabalho propõe uma metodologia para seleção de parâmetros para classificação de faltas do tipo curto-circuito em linhas de transmissão. Uma vez que a maioria dos trabalhos em classificação de eventos em sistemas de potência preocupa-se em classificar um evento de acordo com a morfologia da forma de onda correspondente. Um problema importante e ainda mais difícil é a classificação da causa do evento. Para tal, este artigo utiliza diferentes técnicas de pré-processamento (front-ends), seleção de parâmetros e técnicas de aprendizado de máquina para classificação de faltas do tipo curto-circuito.*

1. Introdução

A crescente exigência dos consumidores de energia elétrica acompanhado ao aumento da produção de equipamentos eletrônicos extremamente sensíveis a distúrbios e oscilações elétricas, aliado à interligação dos sistemas elétricos existentes, assim como, à privatização de concessionárias de energia são fatores que influenciam diretamente em requisitos de Qualidade de Energia Elétrica QEE [1] cada vez mais rigorosos.

Buscando atender a tais requisitos, o setor elétrico dispõe de tecnologias avançadas para a aquisição e armazenamento de informações, as quais inclui os chamados "distúrbios" ou "eventos de interesse" [Bollen et al. 2009]. Um típico exemplo dessa tecnologia são os equipamentos de oscilografia [Luo and Kezunovic 2005] que buscam, com auxílio de algoritmos relativamente simples, anomalias na forma de onda de tensão, assim, se no momento da análise da forma de onda a variação for maior que um determinado valor predefinido, um circuito programável denominado de *trigger* (gatilho),

[1]Segundo [Bollen et al. 2009], é qualquer distúrbio ou alteração nos valores de tensão, corrente ou desvio da frequencia que resulte em falta ou má operação dos equipamentos dos consumidores

inicia o armazenamento dos dados além de informações adicionais como a data e a hora do acorrido.

Apesar da tecnologia avançada, o setor elétrico ainda enfrenta muitas dificuldades para se beneficiar plenamente do processamento automático da informação, tais como algoritmos de mineração de dados. Uma das razões é a falta de base de dados padronizados e livremente disponíveis.

Classificar um determinado evento, ou distúrbio, de acordo com sua morfologia da forma de onda baseando-se, para isso, em parâmetros (tais como *Root Mean Square-RMS*) é um exercício que não apresenta grandes desafios para seus analistas. Uma vez que existem várias normas QEE que classificam os fenômenos de acordo com certas características especificas. Assim, um problema ainda mais desafiador e interessante é a classificação das causas reais deste evento, isto é, se o que provocou o evento de QEE (e.x, afundamento ou elevação de tensão) foi uma descarga atmosférica, um curto-circuito ou um chaveamento de banco de capacitores, por exemplo.

Com base nesse contexto, este trabalho tem como objetivo apresentar uma metodologia para a classificação das causas de eventos de QEE, utilizando para isso técnicas de processamento de sinais e Inteligência Artificia (IA). Para fornecer um exemplo concreto da metodologia proposta, o trabalho concentra-se na análise de faltas do tipo curto-circuito em linhas de transmissão. Esta é uma particular e importante classe de causas de eventos: estudos mostraram que essas faltas foram responsáveis por cerca de 70% dos distúrbios e *blackouts* ocorridos nos sistemas elétricos de potência (SEPs) [Bollen et al. 2009]. Daí a importância do estudo dessa classe de causas de eventos.

2. Classificação de Eventos em Sistemas de Potência

Devido a digitalização dos SEPs, uma crescente quantidade de dados se faz disponível a análises, dos quais boa parte são gráficos com informações referentes as formas de onda (tensão ou corrente) diretamente ligada a eventos de QEE. A classificação destes eventos, seguindo a morfologia de sua forma de onda, é de certa forma, uma tarefa simples. Contudo, uma problemática ainda mais relevante aos especialistas na área é a classificação das causas do distúrbio (evento).

Neste sentido, a distinção entre a "Classificação baseada na morfologia" e a "Classificação da causa do evento" é de extrema importância. Uma vez que a primeira classifica o desvio dos valores nominais da forma de onda de tensão ou corrente e a segunda, as causas reais dos desvios destes valores. A seguir serão caracterizados estes dois tipos de classificação de eventos.

2.1. Classificação de eventos baseada em morfologia

Um Sistema Elétrico de potência típico apresenta três zonas funcionais: geração, transmissão e distribuição. Estas zonas estão sujeitas a distúrbios naturais ou provocados pelo homem que resultam em perturbações que causam variações na forma de onda (morfologia do sinal), caracterizando um evento de QEE.

Em situações normais, a forma de onda de tensão $x(t)$ possui uma frequência especifica que varia entre os valores de 50 ou 60 Hz. Uma vez sabido o valor nominal da amplitude de onda, por exemplo, 500 kV, é conveniente normalizar $x(t)$ por este valor e reportar a amplitude em p.u. (por unidade). Eventos de QEE como, afundamento de tensão,

elevação de tensão e interrupção são definidos em normas internacionais de medições de QEE [IEC 2002] permitindo, assim, a caracterização destes eventos por meio de seus valores de magnitude da forma de onda de tensão (através do RMS, por exemplo) e no tempo de duração do evento.

A classificação automática desses eventos é através de regras de classificação do tipo se-então. Ex: Se RMS $\leq$ 0:01 p.u. e Duração $<$ 1 minuto então o evento é uma interrupção. Em contrapartida, a classificação das causas do evento, apesar de ser mais importante, é menos explorada pela comunidade científica.

2.2. Classificação das causas dos eventos

A classificação das causas dos eventos, por sua vez, é uma tarefa bem mais complexa. Consideremos, por exemplo, que para se diferenciar os eventos causados por queimadas em linhas de transmissão dos eventos causados por descargas elétricas provenientes de raios, devem-se considerar suas "assinaturas" nas formas de onda de tensão e de corrente, visto que se assume, na maior parte dos casos, que não há sensores especiais posicionados ao longo dos sistemas de transmissão, para a captação das características especificas. Outro exemplo mais complexo consiste em distinguir quedas de árvores versus queimadas, caso ambos provoquem um curto na linha de transmissão. Assim, ao escolher as classes, deve-se levar em conta que as causas a serem distinguidas devem imprimir distintas assinaturas nas formas de onda de tensão e corrente. Caso contrário, uma classificação adequada será impraticável.

A classificação de faltas do tipo curto-circuito em linhas de transmissão é um problema de classificação de causa, pois pretende-se descobrir quais as fases estão envolvidas na geração do curto-circuito. Em um contexto mais amplo, poderia-se projetar um sistema de classificação de causas onde "curto-circuito em linha de transmissão" fosse uma classe, e de maneira hierárquica disparar-se o tipo de classificador discutido nesse trabalho para identificar quais as fases envolvidos no curto-circuito detectado pelo módulo superior.

3. Metodologia

A metodologia adotada neste trabalho utiliza a base de dados *UFPAFaults* [Morais et al. 2010]. *UFPAFaults* é uma base de dados pública e devidamente rotulada, desenvolvida no simulador *Alternative Program Program* (ATP) [EMTP 1995]. Modelos ATP têm uma longa história de boa reputação e são muito utilizados para descrever de forma concisa o comportamento atual dos SEPs. Assim, os dados gerados artificialmente para compor a base *UFPAFaults* podem ser (certamente) usados nos experimentos adotados neste trablho.

A metodologia é dividida em duas etapas. Na primeira etapa, simulações experimentais avaliam a performance de vários classificadores de sequência. *Front ends* baseados em Wavelets são comparados com outros mais simples, baseados, por exemplo, nos valores RMS. Esses *front ends* são combinados com diversos algoritmos de aprendizagem: redes neurais artificiais (ANN), árvores de decisão (J4.8), máquinas de vetores de suporte (*Support Vector Machine* - SVM) e K-vizinho mais próximo [Witten and Frank 2005], sugerindo a adoção da arquitetura de classificação de eventos em sistema de potência denominada de Classificação de Sequência Baseada em Quadros (*Frame-Based Sequence Classification* - FBSC) definida em [Morais et al. 2010].

Na segunda etapa os parâmetros dos *Front ends* são concatenados, passando por um processo de seleção de parâmetros (*feature selection*) antes de serem submetidos aos algoritmos de aprendizado de máquina disponíveis no WEKA [2] [Witten and Frank 2005] sobre a base de dados de interesse *UFPAFaults*. Nesta etapa, técnicas de seleção de parâmetros, tais como, filtros [Hall 1999] são investigadas.

4. *Front ends*

Paralelamente a forma de representação que fora adotada, uma simples amostra de dados geralmente não carrega informações suficientementes que permitam o auxílio em tomadas de decisões. Logo, um *Front end* se faz necessário para converter as amostras em parâmetros específicos gerando, assim, as sequências que serão passadas aos algoritmos de classificação.

Todavia, a escolha de um *Front end* específico, o qual apresente um melhor desempenho no processo de classificação de faltas, não é uma tarefa simples. Assim, este trabalho adotou a estratégia que consiste em concatenar as amostras dos *Front ends* em um único conjunto de sequencias. A seguir serão descritos os *Front ends* adotados neste.

4.1. *Front end raw*

Um *front end* é chamado *raw* quando seus parâmetros de saída correspondem a valores das amostras originais, sem qualquer outro processamento que organize as amostras para uma matriz **Z** de acordo com os valores escolhidos de L (número de amostras do quadro) e S (deslocamento do quadro) [Morais et al. 2010]. Por exemplo, consideremos a segmentação de uma falta (ABT) em vetores de parâmetros **z** com 4 quadros. Neste exemplo, $L = 3$ e $S = 1$ o que fornece dois vetores **z**, cada um de dimensão $K = 18$.

4.2. *Front end RMS*

O *Front end* RMS é o mais utilizado permitindo obter uma estimativa aproximada da amplitude da frequência fundamental de uma forma de onda. Este consiste em calcular o valor RMS janelado para cada uma das Q formas de onda. O cálculo do valor do RMS consiste no janelamento de cada uma das formas de onda Q. Considerando o tamanho da janela L e o deslocamento S, o n-ésimo valor RMS $z[n], n = 1, \ldots, N$ de uma forma de onda $x[t], t = 1, \ldots, T$ é dado por

$$z[n] = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(x[l + (n-1)S]\right)^2} \tag{1}$$

4.3. *Front end wavelet*

Quando se adota um *Front end* de multi-resolução tais como wavelets, cuidados especiais devem ser tomados quanto ao seu processamento, devido ao grande número de possibilidades de implementação deste. Caso contrário, a replicação dos resultados pode ser inviável.

[2] Software de mineração de dados de domínio público desenvolvido na Universidade de Waikato na Nova Zelândia. Disponível em http://www.cs.waikato.ac.nz/ml/weka/

Neste trabalho, adota-se dois tipos de *front end* wavelet: waveleconcat que concatena todos os coeficientes wavelet e organiza-os em uma matriz; e waveletenergy que calcula a energia média de cada coeficiente [Morais et al. 2010].

Determinar qual *Front end* se adequa melhor ao processo de classificação de faltas não é uma tarefa simples. Visto isso, a estratégia mais adequada seria concatenar as amostras providas pelos *Front ends* em um único conjunto de sequências.

5. Seleção de Parâmetros

Intuitivamente, quanto mais informações disponíveis ao classificador (parâmetros), melhor seu desempenho. Contudo, isto gera a maldição da dimensionalidade. A medida que cresce o número de parâmetros os dados tendem a ficar exponencialmente esparsos e com isso as amostras aparentam estar igualmente distantes, diminuindo, então, a capacidade de generalização dos classificadores.

A redução da dimensionalidade pode ser realizada, de um modo geral, através de duas abordagens distintas: extração de parâmetros (*feature extraction*), que busca gerar parâmetros mais relevantes ao classificador a partir dos parâmetros existentes. Como exemplo, temos Análise de Componentes Principais *(Principal Components Analysis)* - PCA [Castelli and Bergman 2002] e, a seleção de parâmetros (*feature selection*) que consiste em identificar e reter apenas os parâmetros, dentre o total de subconjuntos possíveis, que mais contribuem para execução de uma dada tarefa. Neste trabalho, adotaremos a abordagem *Feature selection* para redução dos parâmetros.

Os métodos de seleção de parâmetros podem ser classificados basicamente em dois grupos: filtros e *wrappers* [Hall 1999]. Os métodos do tipo filtro avaliam um dado parâmetro através do uso de heurísticas baseadas nas propriedades estatísticas dos dados sendo este independente do classificador adotado. Como consequência, métodos do tipo filtro são geralmente mais rápidos e mais práticos do que os métodos *wrappers* principalmente quando se tem um conjunto de dados de alta dimensionalidade.

6. Experimentos e Resultados

Esta seção descreve os experimentos e os resultados obtidos para a classificação de eventos em sistemas de potência usando a arquitetura FBSC e a base de dados *UFPA-Faults* [Morais et al. 2010]. É assumido ao longo das simulações o uso de um gatilho *(trigger)* [Englert and Stenzel 2002], que descarta as amostras quando a forma de onda não apresenta qualquer tipo de anomalia, e passa aos estágios subsequentes somente o seguimento onde a falta foi detectada.

Foram utilizados os *Front ends* wavelet (waveletconcat, waveletenergy), raw e RMS. A wavelet mãe utilizada foi *Daubechies* 4 com $\gamma = 3$ níveis de decomposição. Consequentemente, para cada uma das $Q = 6$ formas de onda, a decomposição wavelet gerou quatro formas de onda. Os valores de L_{min} e S_{min} foram dentre os seguintes pares: a) $L_{min} = 4$ e $S_{min} = 2$, b) $L_{min} = 9$ e $S_{min} = 4$ e c) $L_{min} = 5$ e $S_{min} = 5$. A escolha por tais valores basearam-se em experimentos anteriores.

Com relação aos *front ends* raw e RMS os valores de L e S foram os mesmos daqueles adotados para os *front ends* wavelet. São apresentados também resultados após a concatenação dos *front ends* e aplicação da seleção de parâmetros baseada em filtros. Os

algoritmos de classificação utilizados foram ANN, J4.8, KNN e SVM, todos pertencentes ao pacote de mineração de dados WEKA.

6.1. Resultados para classificação de sequências

Esta seção apresenta os resultados obtidos para a classificação de sequências E_s usando a arquitetura FBSC e os classificadores ANN, J4.8, KNN e SVM. Foram realizadas diversas simulações em um cenário em que os classificadores foram treinados e testados em sinais livres de ruído. Dos 52 sistemas testados, 17 apresentaram 0% de erro.

As seguintes combinações de *Front ends* e classificadores obtiveram taxa de erro $E_s = 0$ para os três pares de L_{min} e S_{min}: waveletconcat e ANN, waveletconcat e J4.8 e raw com J4.8. Também obtiveram taxas de erros $E_s = 0$ para os valores específicos de $L_{min} = 9$ e $S_{min} = 4$: ANN e waveletenergy, ANN e raw, SVM e waveletconcat, SVM e raw, SVM e RMS. As combinações: ANN e concatfrontend, J.48 e concatfrontend, KNN e raw (para $L_{min} = 4$ e $S_{min} = 2$) também atingiram taxas de erros $E_s = 0$.

Em um cenário com ruído os classificadores apresentaram desempenho variado de acordo com o *Front end* adotado. Os resultados da Figura 1 demonstram que mesmo em um cenário com ruído todos os pares de classificadores, exceto KNN, tiveram uma precisão que permitiram ao classificador de sequencia FBSC alcançar $E_s = 0\%$.

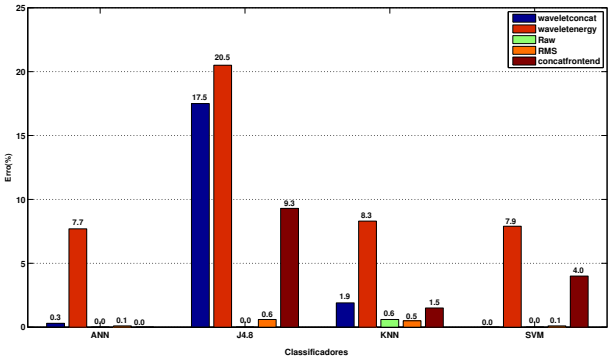
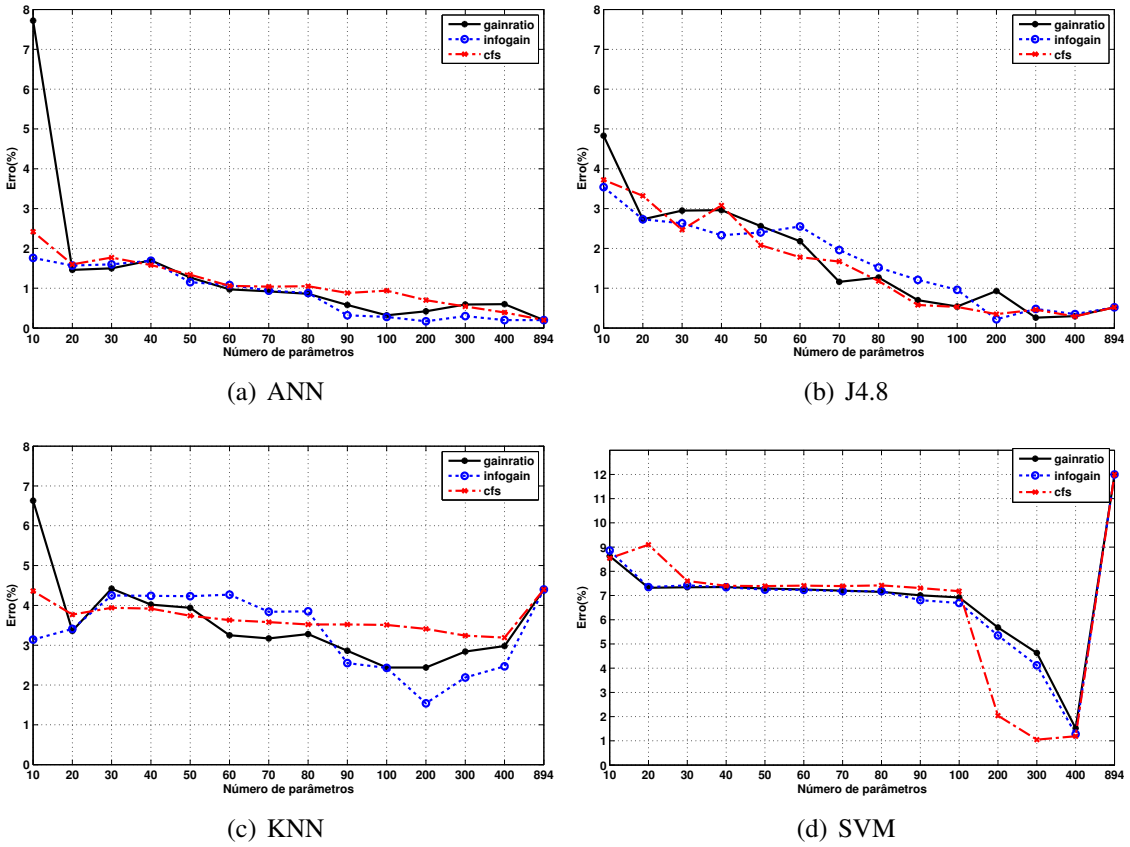


Figura 1. Taxa de erro E_s para a classificação de sequências baseada em quadros usando os *Front ends* waveletconcat, waveletenergy, raw, RMS e a concatenação destes com $L = L_{min} = 9$ e $S = S_{min} = 4$. O conjunto de treinamento foi sem ruído e o de teste teve uma razão de sinal de ruído de 30 dB.

6.2. Resultados com seleção de parâmetros

Para os classificadores ANN e J4.8, com a seleção de 200 parâmetros mais relevantes foi possível obter o mesmo desempenho em relação ao total de parâmetros. Quanto a aplicação da heurística de avaliação, no caso do classificador ANN as menores taxas de erros são obtidas considerando o ganho de informação enquanto que para o J4.8 observa-se um equilíbrio nas taxas de erros obtidas.

Em relação aos classificadores KNN e SVM chama a atenção que com apenas 10 dos 894 parâmetros é possível obter taxas de erros mais baixas. Os resultados descritos são mostrados na Figura 2.



(a) ANN (b) J4.8

(c) KNN (d) SVM

Figura 2. Taxa de erro E_f aplicando seleção de parâmetros nos *front ends* concatenados. Foi utilizado o método de seleção de parâmetros Filtro baseado nas heurísticas de avaliação ganho de informação (*infogain*), razão do ganho(*gainratio*) e correlação (*cfs*) com o número de parâmetros selecionados variando em 10, 20, 30,. . .,100, 200, 300 e 400.

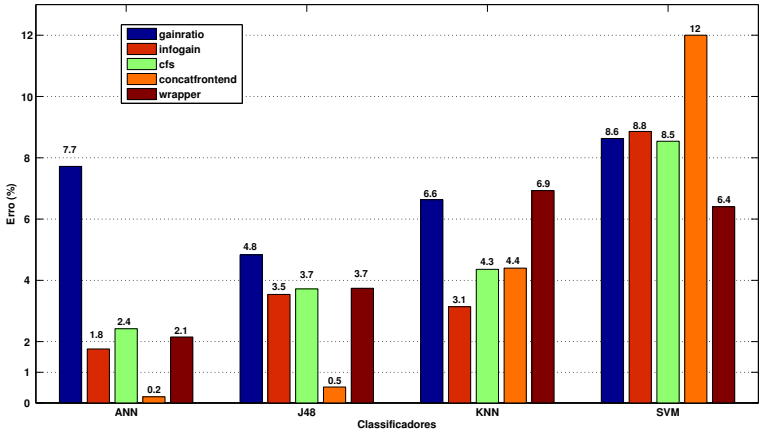


Figura 3. Taxa de erro E_f considerando a concatenação dos *front ends* e aplicando-se as técnicas de seleção de parâmetros filtros e *wrappers* fixando o número de parâmetros em 10.

Observando-se a Figura 3 pode-se inferir que os métodos do tipo filtro possuem desempenho equivalente ou superior em relação a seleção de parâmetros utilizando wrap-

per, exceto para o classificador SVM que apresenta seu melhor desempenho utilizando wrapper.

7. Conclusões

Este trabalho apresentou uma descrição dos aspectos relacionados ao projeto de módulos de classificação de eventos em sistemas de potência. Os resultados mostraram várias relações de compromisso no projeto de um classificador de sequências. A arquitetura FBSC, por exemplo, apresentou um excelente desempenho, visto que as taxas de erro atingiram valores próximos ou iguais a zero mesmo em um cenário com ruído. Outro apontamento, indica que em relação aos classificadores KNN e SVM chama a atenção que com apenas 10 dos 894 parâmetros é possível obter taxas de erros mais baixas. Já os classificadores ANN e J4.8, com a seleção de 200 parâmetros mais relevantes foi possível obter o mesmo desempenho em relação ao total de parâmetros. Por fim, os resultados obtidos neste trabalho permitiram compreender o comportamento dos classificadores na classificação de sequências representando faltas do tipo curto-circuito em linhas de transmissão.

Agradecimentos

Os autores agradecem a Eletrobras-Eletronorte pelo suporte financeiro para o desenvolvimento desta pesquisa.

Referências

Bollen, M., Gu, I., Santoso, S., Mcgranaghan, M., Crossley, P., Ribeiro, M., and Ribeiro, P. (2009). Bridging the gap between signal and power. *Signal Processing Magazine, IEEE*, 26(4):12–31.

Castelli, V. and Bergman, L. (2002). *Image Databases - Search and Retrieval of Digital Imagery*. John Wiley and Sons.

EMTP (1995). *Alternative transient program (ATP) rule book*. Canadian/American EMTP User's Group.

Englert, H. and Stenzel, J. (2002). Automated classification of power quality events using speech recognition techniques. In *14th Power Systems Computation Conference*.

Hall, M. (1999). *Correlation-based feature selection for machine learning*. PhD thesis, University of Waikato.

IEC (2002). Electromagnetic compatibility (EMC) - part 4 - 7: Testing and measurement techniques - general guide on harmonics and interharmonics measurements and instrumentation, for power supply systems and equipment connected thereto.

Luo, X. and Kezunovic, M. (2005). Fault analysis based on integration of digital relay and DFR. *Power Engineering Society General Meeting*, 1:746 – 751.

Morais, J., Pires, Y., Cardoso, C., and Klautau, A. (2010). A framework for evaluating automatic classification of underlying causes of disturbances and its application to short-circuit faults. *Power Delivery, IEEE Transactions on*, 25(4):2083–2094.

Witten, I. and Frank, E. (2005). *Data mining: practical machine learning tools and techniques with java implementations*. Morgan Kaufmann.

Identificação Automática de Serviços a Partir de Processos de Negócio em BPMN

Bruna Christina P. Brandão[1,2], Juliana C. Silva[1] e Leonardo G. Azevedo[1,2,3]

[1] Departamento de Informática Aplicada (DIA)
Universidade Federal do Estado do Rio Janeiro (UNIRIO)
Rio de Janeiro – RJ – Brasil

[2] Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Estado do Rio Janeiro (UNIRIO)
Rio de Janeiro – RJ – Brasil

[3] IBM Research

{bruna.christina, juliana.silva, azevedo}@uniriotec.br, LGA@br.ibm.com

Abstract. *Business Process Models presents several information that shows how enterprises perform their tasks. These models are an important source of information to support service identification in a Service-Oriented Architecture approach, and thereafter, implement process automation. This work implements automatic candidate service identification from business process modeled using BPMN notation. The algorithms implement the heuristics for service identification and consolidation proposed by Azevedo et al. [2009a]. The use of the tool is illustrated by its application in a business process. The candidate service information generated reduces the time required by SOA Analysts during service analysis.*

Resumo. *Modelos de processos de negócio apresentam diversas informações de como processos são executados nas empresas. Em uma Arquitetura Orientada a Serviços, estes modelos são fontes importantes de informações para identificação de serviços com subsequente automatização. Este trabalho implementa a identificação automática de serviços candidatos a partir de processos modelados utilizando a notação BPMN. Os algoritmos foram desenvolvidos a partir de heurísticas de identificação e consolidação de serviços candidatos propostas por Azevedo et al. [2009a]. A ferramenta foi aplicada em um modelo de processo. As informações geradas sobre serviços candidatos reduz o tempo dos Analistas SOA na análise de serviços.*

1. Introdução

Uma das arquiteturas de software que as empresas estão adotando atualmente denomina-se SOA (Arquitetura Orientada a Serviços). SOA é um paradigma para a realização e manutenção de processos de negócio em um grande ambiente de sistemas distribuídos que são controlados por diferentes proprietários [Josuttis, 2007]. SOA traz vantagens para as empresas, tais como, reduções de custos, de riscos, do tempo de desenvolvimento, reutilização de códigos e a possibilidade de um melhor alinhamento com o negócio [Erl, 2005]. O princípio fundamental de SOA é implementar as funcionalidades das aplicações como serviços. Serviços são pedaços de funcionalidades

que possuem interfaces expostas que são invocados via mensagens [Marks *et al.*, 2006].

A identificação de serviços deve ser feita a partir de um processo bem definido e sistematizado [Josuttis, 2007]. A modelagem de processos de negócio é a atividade de representação dos processos de uma empresa de modo que permite que o processo atual seja analisado e melhorado. Um modelo de processos representa várias informações do processo, tais como, atividades, informações de entrada e saída, fluxo de atividades, regras de negócio, requisitos de negócio etc. BPMN (*Business Process Modeling Notation*) é uma notação para modelagem de processo. BPMN foi desenvolvida pela *Business Process Management Initiative* (BPMI) sendo atualmente mantida pelo *Object Management Group* (OMG) [OMG, 2011]. A BPMN foi apoiada por várias empresas mundialmente conhecidas tornando-se a notação mais utilizada [Valle *et al.*, 2012]. Azevedo *et al.* [2009a] propuseram um conjunto de heurísticas que sistematiza a identificação de serviços a partir de modelos de processos de negócio.

O objetivo deste trabalho é automatizar a execução das heurísticas propostas por Azevedo *et al.* [2009a] para gerar as informações sobre serviços candidatos a partir de modelos de processos modelados utilizando a notação BPMN. Azevedo *et al.* [2009b] implementaram a identificação automática de serviços a partir de modelos de processos modelados usando a notação EPC [Scheer, 2000]. Este trabalho implementa as heurísticas para processos modelados com a notação BPMN bem como implementa as heurísticas para tratar controle de fluxo AND, XOR, OR que não haviam sido implementadas por Azevedo *et al.* [2009b].

O restante deste trabalho está dividido da seguinte forma. A Seção 2 apresenta as heurísticas para identificação de serviços. A Seção 3 apresenta como as heurísticas foram automatizadas. A Seção 4 ilustra o uso da ferramenta implementada em um modelo de processo de negócio. Finalmente, a Seção 5 conclui o trabalho.

2. Identificação de serviços

Serviço candidato é uma abstração (não implementada) de serviço a qual, durante a fase de projeto em um modelo de ciclo de vida, pode ser escolhida para ser implementada como um serviço ou como uma funcionalidade de uma aplicação [Erl, 2005].

Em modelos de processos de negócio, serviços candidatos podem ser identificados a partir de informações estruturais (por exemplo, fluxo do processo) e semânticas (por exemplo, informações de entrada e saída, regra de negócio e requisitos do negócio) [Azevedo *et al.*, 2009a]. O método de identificação de serviços possui três etapas [Azevedo *et al.*, 2009a]: (i) seleção de atividades; (ii) identificação e classificação de serviços candidatos; (iii) consolidação de serviços candidatos. Na etapa de "Seleção de atividades" selecionam-se as atividades do modelo de processo que serão utilizadas para identificação de serviços. Na segunda etapa, heurísticas são aplicadas para identificação dos serviços candidatos. Na terceira etapa, heurísticas são utilizadas para consolidar informações sobre os serviços, como, por exemplo, eliminar serviços iguais ou para juntar serviços que contenham funções semelhantes.

2.1. Heurísticas para identificação de serviços candidatos

Esta seção apresentada as heurísticas definidas por Azevedo *et al*. [2009a] e como as informações necessárias consumidas pelas heurísticas são modeladas.

Em BPMN, as regras de negócio são modeladas no diagrama de processo, através de atividades marcadas com o símbolo [OMG, 2011], no qual corresponde a marcação de um atributo no objeto que representa este elemento. A definição da heurística de regra é a seguinte: "Toda regra de negócio deve ser identificada como serviço candidato". Logo, para a implementação desta heurística, são geradas informações sobre serviços candidatos para todos os elementos marcados com este símbolo. Porém, não existe símbolo para representar requisito de negócio [Pavlovski *et al.*, 2008] e para representar interfaces de processo [OMG, 2011], logo, não foi possível automatizar as heurísticas para estes elementos.

As informações de entrada e de saída das atividades são representadas em BPMN pelo objeto de dados (*data object*), cujo símbolo é , e portadores de informação (*data store*) são representados pelo símbolo . Logo, para a heurística de identificação de serviços candidatos a partir de informações de entrada e saída ("Toda informação de entrada ou saída de uma atividade associada a um portador de informação deve ser considerada como um serviço candidato."), gera-se informações de serviços candidatos para todos os objetos de dados do modelo de processo associados à um *data store*.

A Figura 1 apresenta um exemplo de parte de um modelo de processo. A partir deste modelo, como exemplo, tem-se que são identificados serviços candidatos: para a atividade "Determinar taxas de juros a ser cobrada do cliente" que representa uma regra de negócio; e para os objetos de dados "Taxa de juros" e "Taxa de juros do cliente", neste caso, o primeiro corresponde a um serviço de leitura de dados enquanto que o segundo é um serviço de escrita de dados.

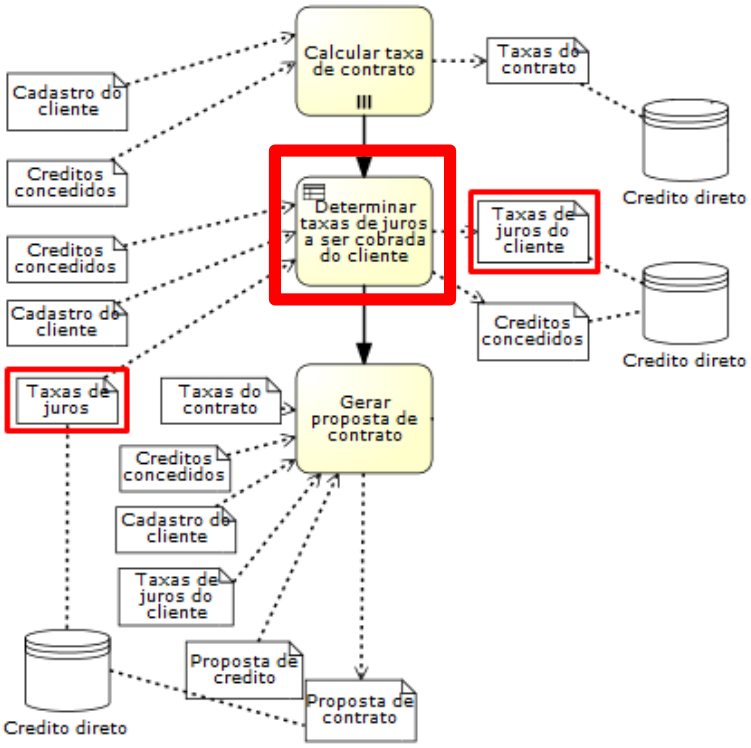


Figura 1 - Exemplo de regras de negócio e informações de entrada e saída.

Na notação BPMN a atividade de múltipla instancia é modelada no diagrama de processo através de atividade marcada com o símbolo [OMG, 2011]. Logo, a partir

dos elementos do modelo de processo marcados com este símbolo, serviços candidatos são identificados de acordo com a heurística “Toda atividade de múltipla instância deve ser considerada um serviço candidato”. No exemplo da Figura 2, um serviço candidato é identificado para a atividade “Calcular taxa de contrato”.

A heurística de identificação de serviços candidatos a partir de atividades sequenciais (“Toda sequência de duas ou mais atividades identificadas no processo deve ser considerada um serviço candidato.”) gera automaticamente informações a partir da identificação de a sequência de atividades da Figura 2.

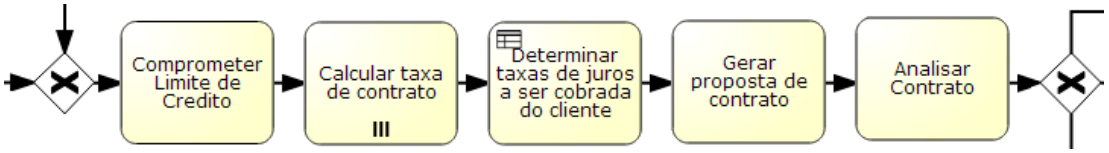


Figura 2 - Exemplo de atividades sequenciais.

As heurísticas de identificação de serviços candidatos a partir de workflow AND, XOR e OR são semelhantes. No caso da heurística de AND (“Todo fluxo paralelo deve ser considerado como um serviço candidato, incluindo o seu início na bifurcação até a sua junção ou término em evento(s) final(is).”), o workflow AND corresponde a todos os elementos existentes entre o gateway AND ⊕ de abertura e o de fechamento. Estes gateways são responsáveis por controlar a execução das atividades em paralelo. O fechamento deste fluxo também poderia ter ocorrido através de um evento final ○. Um serviço é identificado a partir dos elementos contidos entre estes dois elementos. A Figura 3 apresenta um exemplo de modelo a partir do qual um serviço candidato é identificado. A identificação corresponde a identificar estes elementos de *gateway* AND de abertura e o seu fechamento ou evento final, e gerar o serviço candidato composto pelo conjunto de atividades que se encontram dentre estes elementos.

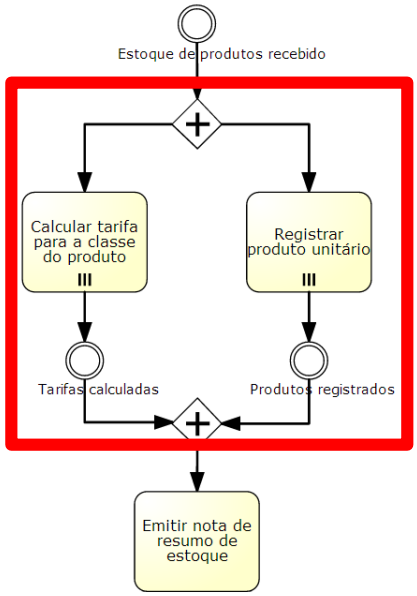


Figura 3 - Exemplo de workflow AND com fluxos paralelos sincronizados.

O loop de atividades corresponde às atividades que repetem no fluxo devido ao retorno de uma atividade para o início da sequência. A identificação de serviços

candidatos para a heurística de LOOP (“Toda estrutura de workflow onde duas ou mais atividades são executadas repetidas vezes deve ser considerada um serviço candidato”) corresponde identificar a repetição de atividades no processo e a gerar o serviço candidato para as mesmas, ou seja, foi identificado um serviço candidato a partir de loop de atividades.

2.2. Heurísticas de consolidação

Azevedo *et al*. [2009a] definiram as seguintes heurísticas de consolidação de serviços candidatos: heurística de eliminação de serviços duplicados; heurística de associação de serviços com papéis; heurística de associação de serviços com atividades; heurística de geração de informações sobre de serviços identificados a partir de fluxo (número de raias envolvidas no fluxo e números de subfluxos).

A heurística de consolidação de eliminação de serviços duplicados (“Deve ser mantido apenas um serviço quando existirem serviços candidatos duplicados”) é utilizada quando existem dois ou mais serviços com nomes iguais e mesma origem. A heurística de consolidação de serviços com papéis (“Todo serviço candidato deve ser associado aos papéis (raias) em que suas atividades estão localizadas”) é essencial para o gerenciamento e restrição de acesso as funcionalidades do processo. Esta heurística foi implementada observando os papéis responsáveis por atividades a partir da existência de elementos associados dos quais foram identificados serviços. A heurística de associação de serviços com atividades (“Todo serviço candidato deve ser associado às atividades a partir do qual foi identificado”) lista para os serviços as atividades associadas a elementos a partir dos quais serviços foram identificados. Por exemplo, na heurística de loop, o serviço deve ser associado com as atividades presentes no loop. A heurística de consolidação de serviços identificados a partir de fluxo, por exemplo, contabiliza o número de raias que contiver atividades presentes nos serviços candidatos de fluxo e o número de subfluxos existentes em um fluxo.

3. Automatização da identificação

Todas as heurísticas foram implementadas considerando algumas premissas: todas as atividades devem possuir nomes distintos, e de acordo com a especificação da notação BPMN, o elemento atividade só pode possuir uma transição de entrada e uma transição de saída.

Para a identificação automática dos serviços definidos pelas heurísticas foi desenvolvida uma aplicação em Java, no qual lê um processo no formato JSON. Em seguida, o sistema lê esse arquivo e mapeia os elementos existentes no JSON para a estrutura de *hashmaps*. Um *hashmap* é criado para cada tipo de elemento do modelo de processo (atividades, gateways, eventos e outros). Posteriormente, é iniciado o processamento das heurísticas de identificação.

Para a heurística de regras de negócio, é percorrido todo o *hashmap* de atividade e identificada as atividades que são regras de negócios. Essas atividades são consideradas serviços candidatos. Para a heurística de múltipla instância, o procedimento é parecido com o anterior. O *hashmap* de atividades é percorrido e, para cada elemento, é verificado se é uma atividade de múltipla instância. Caso positivo, essas atividades também são consideradas serviços candidatos. Para a identificação da heurística de informações de entrada e saída, o *hashmap* de objetos de dados é

percorrido e é verificado se este está ligado a algum *data store*. Caso positivo, esses elementos são considerados serviços candidatos. Para atividades sequenciais é necessário armazenar os dados em uma estrutura de lista encadeada, o que possibilita visualizar o elemento e todos os demais elementos que estão diretamente ligados a ele. Neste caso, para cada atividade ligada à outra atividade, se houver uma sequência de duas ou mais atividades seguidas, então ela é considerada um serviço candidato. As demais heurísticas de identificação de serviços são heurísticas de workflow.

Para tratar esses casos, foi necessário armazenar os dados na estrutura de RPST (*Refined Process Structure Tree*). RPST (*Refined Process Structure Tree*) é uma técnica de análise de fluxos que representa um modelo de processo em uma árvore hierárquica, possibilitando uma navegação mais simples para acessar partes do modelo para executar algoritmos [Vanhatalo *et al.*, 2009]. RPST é baseado no fato de que cada grafo pode ser decomposto em uma hierarquia de subgrafos independentes de forma lógica, contendo uma entrada e uma saída. A hierarquia é mostrada como uma árvore onde a raiz é a própria árvore e as folhas são fragmentos. Vanhatalo *et al.* [2009] classificam esses fragmentos como triviais (T), *bonds* (B), polígonos (P) e rígidos (R). Fragmentos triviais correspondem a dois nós conectados por um único arco. Os *bonds* são conjuntos de fragmentos que compartilham dois nós em comum. Nos processos em BPMN, os fragmentos *bonds* normalmente são resultantes dos gateways de bifurcação e de junção. Polígonos correspondem a sequências de outros fragmentos. Fragmentos que não se encontram nas classificações dos fragmentos triviais, *bonds* ou polígonos são classificados como rígidos [Polyvyanny *et al.*, 2011].

As informações geradas durante o processamento são armazenadas em arquivos que o Analista SOA utilizar para fazer usas análises.

4. Aplicação das heurísticas

A avaliação da proposta de automatização da identificação de serviços candidatos foi realizada executando a ferramenta implementada no modelo de processo "Analisar pedido de crédito". Este modelo de processo é apresentado de forma compacta (sem os *data objects* e *data stores*) na Figura 4. A interface da ferramenta está disponível em http://uniriotec.br/~azevedo/soa_bpm/AplicacaoUI.png. O modelo completo está disponível em http://uniriotec.br/~azevedo/soa_bpm/AnalisarPedidoDeCredito-BPMN.png. Este processo possui um evento inicial, quatro eventos finais, dezoito atividades, quinze *data stores*, treze eventos intermediários, quarenta e nove *data objects* e seis *gateways*. O resultado da identificação dos serviços está disponível em http://uniriotec.br/~azevedo/soa_bpm/Services.xls e a consolidação das informações dos serviços em http://uniriotec.br/~azevedo/soa_bpm/ServicesConsolidated.xls.

No processo modelado é feita a análise de uma proposta de crédito recebida, verificando se crédito pode ser concedido ao cliente requisitante. Primeiramente, ao receber uma proposta de crédito, é verificado o cadastro do cliente. Caso este cliente não possua cadastro, deverá ser realizado o cadastramento. Caso contrário, se as informações do cliente estiverem desatualizadas, apenas deverá ser feito uma atualização das informações. Caso contrário, segue-se para o próximo passo. Posteriormente, é verificado o limite de crédito deste cliente. Caso o limite não seja aprovado, a proposta de crédito é cancelada. Se o limite de crédito for aprovado, deve-se comprometer o limite de crédito, calcular a taxa de contrato e de juros e gerar a proposta de contrato. O analista de crédito deve analisar o contrato e realizar ajustes

necessários. Ele poderá cancelar o contrato caso o considere um contrato de risco. Em seguida, o cliente deverá verificar as condições do contrato e cancelar ou aprovar o contrato, finalizando o processo.

Neste processo foram identificados dois serviços candidatos de regras de negócio, um serviço candidato de múltipla instância, dezenove serviços candidatos de entrada ou saída, três serviços candidatos de atividades sequenciais, um serviço candidato de *loop* e quatro serviços candidatos de workflow XOR, totalizando 30 serviços. A ferramenta levou 5 segundos para executar as heurísticas e gerar os arquivos de saída. Os arquivos são gerados no formato Excel: um arquivo para as informações dos serviços identificados a partir das heurísticas de identificação; e, outro arquivo para as informações consolidadas.

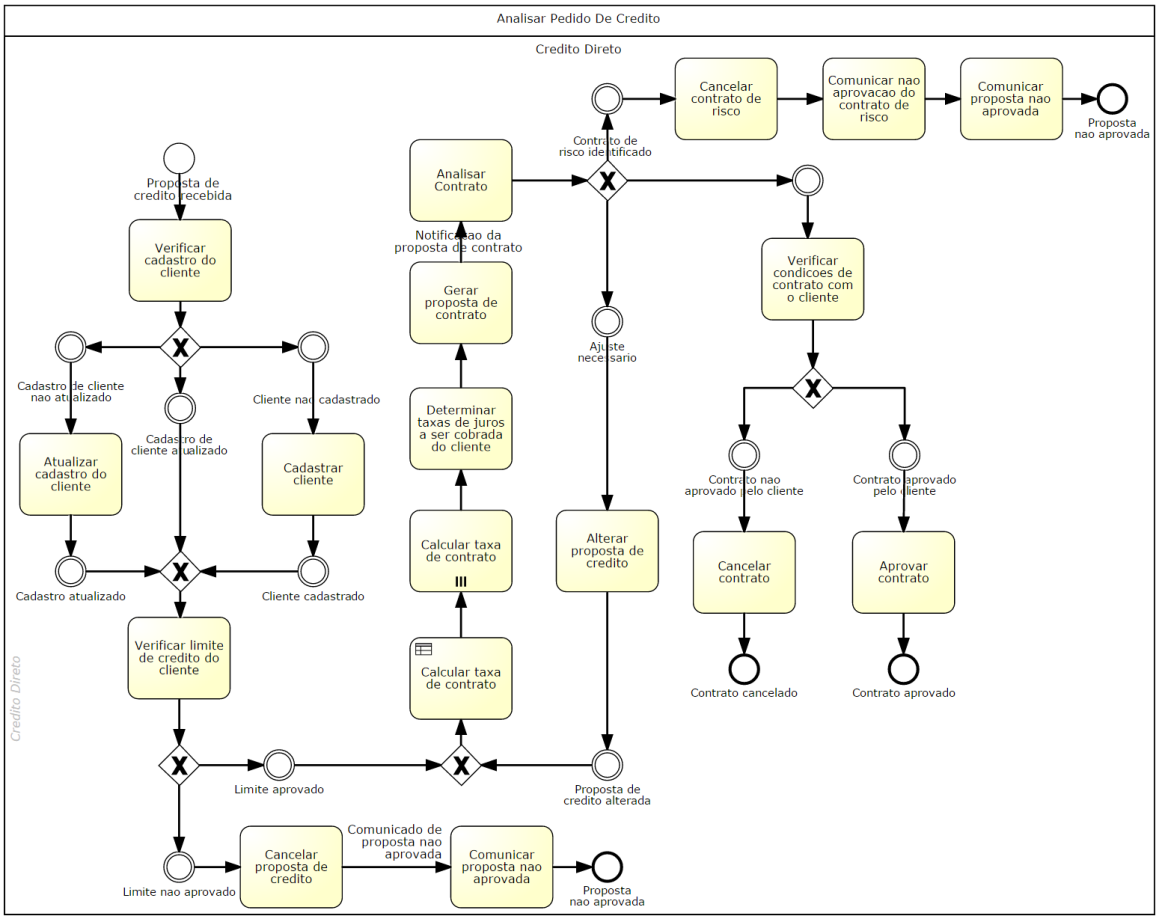


Figura 4 - Processo Analisar de Pedido de Crédito em BPMN [Adaptação de Sousa *et al.*, 2012].

5. Conclusão

Este trabalho implementou as heurísticas para identificação de serviços propostas por Azevedo *et al.* [2009a]. O benefício mais claramente identificado com a execução desta automatização foi a velocidade com que os serviços foram identificados, reduzindo riscos de erros humanos. Esta tarefa, quando executada manualmente, demanda tempo e esforço que foram poupados com a execução da ferramenta desenvolvida, no qual demorou segundos para executar. Outro benefício identificado foi a possibilidade de que qualquer alteração na modelagem do processo não demande muito tempo para realizar novamente a identificação dos serviços. Ou seja, a execução da ferramenta permite

gerar rapidamente novos serviços, de acordo com os ajustes realizados.

Como limitações do trabalho, duas heurísticas desenvolvidas por Azevedo *et al.*, [2009a] não puderam ser implementadas devido às limitações da notação BPMN: heurística de identificação de serviços por requisitos de negócio; e heurística de identificação de interface de processo. Outra limitação deste trabalho aparece na heurística de consolidação de serviços, na qual foi o fato da identificação ter sido executada no arquivo correspondente ao modelo de processo e não em um repositório de processos. Dessa forma, não foi possível calcular o grau de reuso de cada serviço identificado.

Referências

Azevedo, L. G., Santoro F., Baiao F., Souza J. F., Revoredo K., Pereira V., Herlain I. (2009a). "A Method for Service Identification from Business Process Models in a SOA Approach". In: 10th Workshop on Business Process Modeling, Development, and Support (BMDS'09), v. 29. p. 99-112.

Azevedo L. G., Sousa H. P., Souza J. F., Santoro F., Baiao, F. (2009b) "Identificação automática de serviços candidatos a partir de modelos de processos de negócio." Conferência IADIS Ibero Americana WWW/INTERNET 2009 (CIAWI'09).

Erl T. (2005) Service-Oriented Architecture (SOA): Concepts, Technology, and Design. Upper Saddle River, NJ, USA: Prentice Hall.

Josuttis, N. (2007) SOA in practice: The Art of Distributed System Design. Beijing; Cambridge. O'Reilly, 324p.

Marks E. A., Bell, M. (2006) Service Oriented Architecture (SOA): a planning and implementation guide for business and technology. John Wiley & Sons.

OMG. (2011) "Business Process Model and Notation.".

Pavlovski C. J., ZOU J. (2008) "Non-functional requirements in business process modeling." In APCCM '08: Proceedings of the fifth on Asia-Pacific conference on Conceptual modelling, 2008.

Polyvyanny A., Vanhatalo J., Völzer H. V. (2011) "Simplified Computation and Generalization of the Refined Process Structure Tree.". Web Services and Formal Methods. Springer Berlin Heidelberg. 25-41.

Scheer, A.-W. (2000), ARIS - Business Process Modelling. Springer, Berlin, Alemanha.

Vanhatalo J., Völzer H. V., Koehler J.(2009) "The refined process structure tree.". Data & Knowledge Engineering, v. 68, n. 9, p. 793-818.

Valle R., Oliveira S. B. (2012) "Análise e Modelagem de Processos de Negócio. Foco na Notação BPMN." São Paulo: Atlas.

Avaliação de Aspectos de Usabilidade em Ferramentas para Mineração de Dados

Clodis Boscarioli[1], José Viterbo[2], Mateus Felipe Teixeira [2]

[1]Curso de Ciência da Computação – Universidade Estadual do Oeste do Paraná (UNIOESTE) - Cascavel, Paraná, Brasil

[2]Instituto de Computação - Universidade Federal Fluminense (UFF)- Niterói, Rio de Janeiro, Brasil

clodis.boscarioli@unioeste.br, {viterbo,mateus.teixeira}@ic.uff.br

Abstract. *Currently various techniques and tools are proposed to allow the end user to interpret large volumes of data stored in organizational databases for a given decision taking. In this paper, we discuss the ways to interact with the tools of cluster analysis, focusing on both the grouping as the stages of interpretation. Investigated how the usability and user experience in such tools can improve the understanding of the discovered knowledge and thus improved decision making made on a database. We evaluated four different cluster analysis tools: Knime, Orange Canvas, Rapidminer Studio and Weka data mining tools.*

Resumo. *Atualmente várias técnicas e ferramentas são propostas para permitir que o usuário final possa interpretar grandes volumes de dados armazenados em bancos de dados organizacionais para uma determinada toma de decisão. Neste trabalho, discutimos as formas de interagir com as ferramentas de análise de cluster, tendo em conta tanto o agrupamento quanto as etapas de interpretação. Investigamos como a usabilidade e a experiência do usuário em tais ferramentas pode melhorar a compreensão do conhecimento descoberto e assim melhoras a tomada de decisão feita sobre uma base de dados. Foram avaliados quatro ferramentas de análise de agrupamento diferentes: Knime, Orange Canvas, Rapidminer Studio e Weka ferramentas de mineração de dados.*

1. Introdução

Devido ao rápido crescimento do volume de dados armazenados em bancos de dados organizacionais e às limitações humanas na análise e interpretação de dados, técnicas apropriadas são necessárias para permitir a identificação de uma grande quantidade de informação e conhecimento em tais bancos de dados. Esse processo analítico é conhecido como *Knowledge Discovery in Databases* (KDD), um processo não trivial que visa descobrir novos padrões úteis e acessíveis em bancos de dados [Fayyad, 1996], e compreende três etapas principais: o pré-processamento para a preparação de dados, a mineração de dados e o pós-processamento, que inclui a depuração e/ou síntese dos padrões descobertos.

A mineração de dados (*data mining*) é o núcleo do processo de KDD e se baseia em um conjunto de diferentes algoritmos para extrair padrões ocultos em bases de dados, que variam de acordo com o objetivo da análise, que pode ser, entre outros, a

definição de modelos de regressão, classificação ou agrupamento de dados. Agrupamento de dados, de interesse particular neste estudo, pode ser definido como a identificação de grupos ou subconjuntos de dados em que há uma coesão interna elevada entre os objetos que pertencem a um grupo, mas também um grande isolamento externo entre os grupos.

Assume-se aqui que processo de agrupamento de dados é composto não apenas pela aplicação parametrizada de algoritmos para identificação de grupos, mas também por uma segunda etapa, que consiste em realizar a interpretação dos resultados gerados pelos algoritmos de aprendizado não supervisionado, o que pode ser feito pela aplicação de métodos de visualização de dados. O principal objetivo da visualização de dados é integrar o usuário final no processo de descoberta de conhecimento, proporcionando uma representação gráfica dos dados originais, ou do resultado de agrupamento de dados.

Neste trabalho, investigamos como os aspectos de usabilidade e a experiência do usuário na interação com ferramentas de análise de agrupamento podem melhorar a compreensão do conhecimento descoberto. Para este fim, avaliamos quatro ferramentas de análise de agrupamento gratuitas e amplamente utilizadas: Knime, Orange Canvas, Rapidminer Studio e Weka.

Na próxima seção, apresentamos alguns conceitos básicos sobre o agrupamento de dados e visualização de dados. Na Seção 3, descrevemos as ferramentas selecionadas para o nosso estudo. Na Seção 4, descrevemos a metodologia de avaliação de usabilidade de adotada. Na Seção 5, discutimos os resultados observados. E, finalmente na Seção 6, apresentamos algumas conclusões e perspectivas da pesquisa.

2. Mineração e Visualização de Dados: Conceitos Introdutórios

Agrupamento de dados é uma denominação geral para métodos computacionais que analisam dados visando descoberta de conjuntos de observações homogêneas [Everitt, 2001]. Considerando uma base de dados com n padrões, cada um destes medido segundo p variáveis, o objetivo é encontrar uma relação que os agrupe estes padrões em k diferentes grupos. Com isto espera-se a visualização da relação entre os dados que anteriormente à operação do agrupamento de dados não era explícita.

Existem vários algoritmos para agrupamento de dados, que utilizam de maneiras diferentes para a identificação e a representação dos resultados do algoritmo. A escolha de cada algoritmo depende do tipo de dado que se pretende explorar, da aplicação e objetivos de análise. Neste trabalho, optamos por utilizar apenas o algoritmo *k-means* [Macqueen, 1967], por estar disponível em todas as ferramentas analisadas e ser um dos mais simples e mais utilizados para a tarefa de agrupamento.

No *k-means*, inicialmente define-se o número de grupos k a serem identificados. Em seguida, são definidos os k centroides iniciais, podendo-se utilizar diversas heurísticas. Para a etapa seguinte cada item da base de dados é associado ao centroide mais próximo, de acordo com uma medida de distância pré-estabelecida e ao final, recalculam-se os k centroides calculando a média para cada atributo entre os itens associados ao centroide analisado no momento. Este processo é repetido até que não haja mais mudança nos grupos formados.

A ideia principal da visualização de dados é integrar o usuário final ao processo, oferecendo-lhe uma representação gráfica dos dados. Técnicas de visualização são utilizadas após a mineração de dados para permitir ao usuário final interpretar os resultados de um determinado algoritmo. No caso de agrupamento de dados, por exemplo, atua identificando características de um determinado grupo, relação entre padrões e distinções entre grupos, distribuição espacial de padrões, entre outros. O usuário também poderá interagir com a representação gráfica para interpretar de uma melhor maneira o resultado gerado pelo método aplicado [Wong, 1999].

Há diferentes técnicas de visualização de dados, porém, a grande maioria é limitada pela dimensionalidade da base de dados a ser explorada. Com o passar do tempo, várias técnicas foram desenvolvidas para diferentes tipos de dados e também para diferentes dimensionalidades.

3. As Ferramentas Analisadas

As ferramentas selecionadas para avaliação neste trabalho são brevemente descritas a seguir. Para cada uma, primeiramente foram identificadas as técnicas de agrupamento de dados disponíveis e também as técnicas de visualização aplicáveis a agrupamento de dados.

O Knime [Knime, 2013] é uma ferramenta proposta para o uso na mineração de dados, estatística e outras áreas. Possui diversos métodos que possibilitam uma extração de conhecimento completa de uma determinada base de dados. O funcionamento da Knime é todo baseado na ideia de adição de nodos de métodos a um fluxo de execução. Porém a ferramenta apresenta algumas funcionalidades ainda não explicitadas, o que pode dificultar o seu uso.

Para o *k-means*, traz como resultado textual somente a formação dos centroides, apresentando os valores que cada atributo foi composto, e por quantos padrões este centroide foi constituído.

O Orange Canvas [Canvas, 2013] é uma ferramenta open source de mineração de dados com enfoque para classificação de dados, regressão de dados e mineração visual de dados, mas também possui métodos de associação de dados. O fluxo de execução da ferramenta é simples, e como a Knime apresenta a estrutura de nodos em que cada nodo adicionado ao campo de fluxo irá executar uma determinada tarefa. Também apresenta um sistema de *feedback* para cada método, que retorna ao usuário os dados de entrada e de saída de cada método.

O Rapidminer Studio [Rapidminer, 2013] é uma ferramenta proprietária, que possui uma versão de testes, de mineração de dados também voltada à estatística, banco de dados e processos de análises em dados. Tem como o principal foco disponibilizar um ambiente de trabalho totalmente gráfico, com a presença de elementos gráficos que significam uma operação em questão, por exemplo, um método de mineração de dados. A ferramenta possui um fluxo de execução intuitivo, seguindo a ideia de fluxo de execução baseada em nodos que representam um determinado processo. O retorno é textual, apresentando o número de grupos formados e por quantos padrões estes são formados, e também uma tabela de dados.

O Weka [Weka, 2013] é a uma ferramenta de aprendizado de máquina aberta usada para tarefas de mineração de dados, contém técnicas para pré-processamento de

dados, classificação, regressão, agrupamento, regras de associação e visualização. Como resultado da execução do método *k-means*, Weka apresenta o número de iteração, a soma do quadrado do erro dentro dos grupos e os centroides formados. A técnica de gráfico de dispersão da Weka não mostra a dispersão dos dados de um determinado grupo, e sim a dispersão dos dados de acordo com um de seus atributos em função do grupo que esta pertence.

4. Metodologia de Avaliação

Na avaliação das ferramentas realizou-se a inspeção por meio do Percurso Cognitivo, um método de avaliação de IHC por inspeção cujo principal objetivo é avaliar a facilidade de aprendizado de um sistema interativo, através da exploração de sua interface [Whartonm, 1994]. Além disso, testes de usabilidade foram realizados. Para este fim, foram definidos dois grupos de usuários, profissionais e estudantes da área das ciências exatas com algum contato prévio com a área de mineração de dados, que após a realização de uma atividade proposta em um cenário real, com duração de tempo necessária para cada um percorrer o percurso e as tarefas dadas, responderam um questionário contendo questões sobre o perfil do usuário e da usabilidade da ferramenta.

As seguintes tarefas foram definidas para serem executadas em todas as ferramentas:

1. Carregar arquivo de dados "Iris.arff";
2. Escolher a Tarefa de Agrupamento de Dados – método *k-means* com $k = 3$. Para os demais parâmetros foram utilizados os valores default;
3. Executar o algoritmo;
4. Visualizar os resultados.

4.1 Percurso Cognitivo

Embora os percursos para realizar as ações sejam diferentes em cada ferramenta, pode-se concluir que o usuário vai tentar atingir o resultado correto desde que ele saiba o funcionamento teórico dos métodos implementados para a realização de ajustes dos parâmetros, bem como o conteúdo da base de dados a ser analisada para verificação semântica dos resultados obtidos.

Para tal, as seguintes ações deveriam também ser realizadas:

- Identificar botões que possibilitem carregar a base de dados e especificar o caminho;
- Identificar botões que possibilitem inserir o método *k-means* e configurar os parâmetros de entrada;
- Identificar botões que possibilitem inserir métodos de visualização e interpretar resultado.

Ademais, há particularidades de interação entre elas. O usuário necessita saber em que categoria um determinado item é classificado para poder encontrá-lo nas ferramentas. No Weka, para a ação de abrir/carregar a base de dados pode acontecer

erros na identificação do componente da ferramenta. Por exemplo, em alguns pontos o componente chama-se “ARFF Reader” e noutros, é disponibilizado como “FILE”, ficando implícito que este componente também poderá abrir/carregar um arquivo de extensão ".arff". Caso não saiba a priori, o usuário deverá aprender enquanto realiza a tarefa qual é o fluxo necessário para realização das atividades de mineração de dados almejadas, em cada uma das ferramentas.

5. Análise de Resultados

Conforme ilustrado na Figura 1, pôde-se observar que o grupo de professores, num total de quatro, possuía maior experiência com as ferramentas que os oito alunos que participaram do experimento. Além disso, os gráficos mostram que a maioria dos usuários está bem familiarizada com Weka, enquanto muitos nunca usaram Knime ou Orange Canvas.

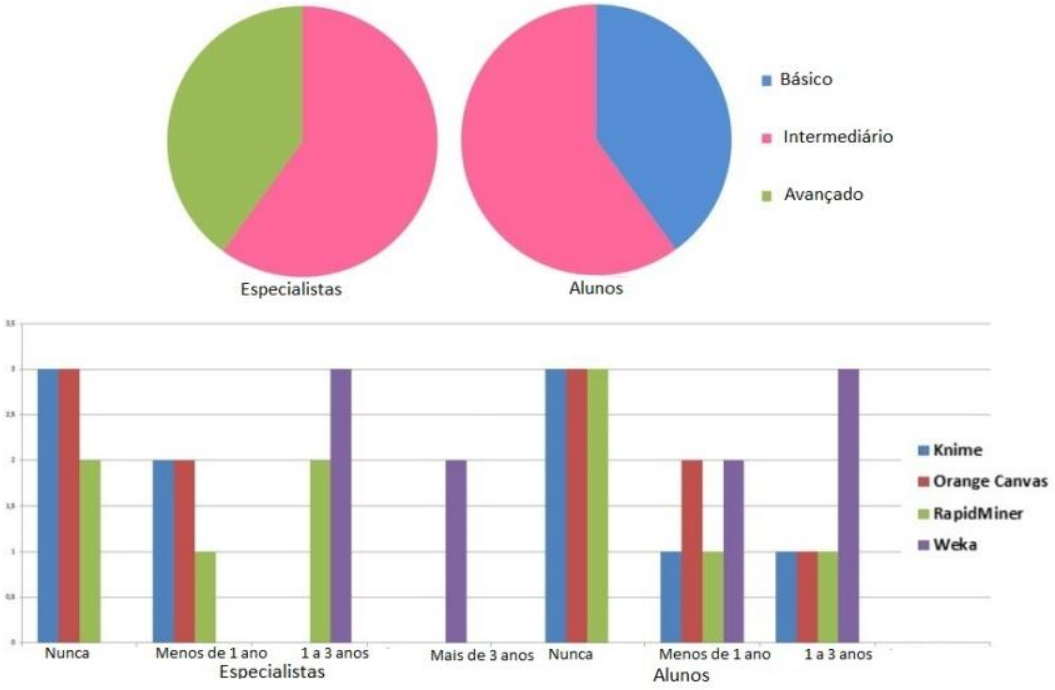


Figura 1. Nível de expertise (acima) e familiaridade com as ferramentas (abaixo) de dois grupos diferentes de usuários.

A Figura 2 apresenta os resultados da avaliação das ferramentas pelos usuários. As ferramentas Knime, Rapidminer Studio e Orange Canvas deixam o usuário interagir mesmo quando uma ação está sendo feita de forma errada, avisando-o do erro somente quando se tenta executar o fluxo do processo de mineração de dados. Este comportamento faz com que o usuário perca tempo na sua interação, podendo fazer com que o fluxo seja refeito, por não saber onde exatamente ocorreu o erro. Uma solução para isto seria o monitoramento das ações do usuário e a notificação imediata de um possível erro.

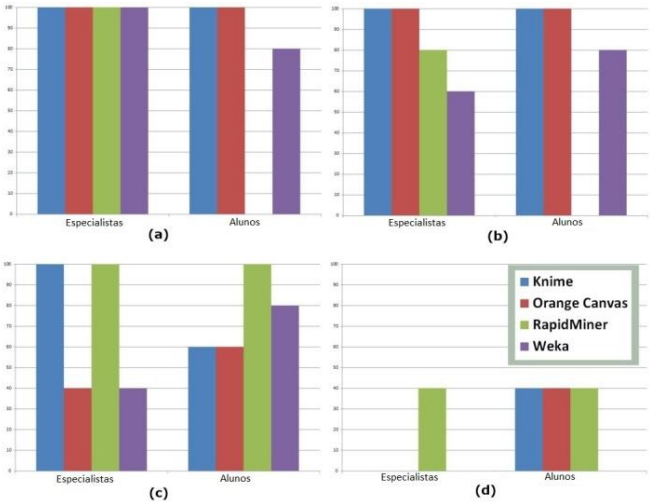


Figura 2. Proporção de usuários que avaliaram positivamente (a) a distribuição de informações, (b) os ícones e comandos na interface, (c) a descrição das tarefas disponíveis nas ferramentas e (d) a disponibilidade de mensagens de erros.

A respeito da interface, o grupo dos alunos indicou as telas carregadas de informações, já o grupo de professores não. Foi citado que Weka tem funções que não são acessadas de forma intuitiva. De acordo com os usuários, as ferramentas dispõem de informações bem distribuídas e destacadas, o que contribui para o aprendizado. Somente Rapidminer Studio, quando avaliada pelos alunos, não obteve boa avaliação. Isso se dá quando o conjunto de objetos necessário para execução de uma tarefa é grande e complexo, fazendo com que seja necessário um conhecimento prévio das suas funcionalidades, conhecimento já presente no grupo dos professores.

As ferramentas foram mal avaliadas no que diz respeito à objetividade dos ícones e botões disponíveis nas telas, ou seja, há elementos que atrapalham e/ou dificultam a escolha de uma determinada ação. Obtiveram má avalição também no que diz respeito das descrições dos elementos contidos na interface e nas opções de ajuda *online*.

Os resultados em relação à facilidade da utilização de cada ferramenta podem ser observados na Figura 3. De acordo com os estudantes, a Orange Canvas é que apresenta a interface onde é mais fácil encontrar as informações e elementos desejados e a Rapidminer Studio foi a pior avaliada. Já os professores, mostraram um resultado mais homogêneo entre as ferramentas, pois já houve um contato com uma ou mais ferramentas.

já que para a execução todas as ferramentas apresentam um fluxo de execução bastante claro, que tem um início, meio e fim.

Acreditamos que a integração interdisciplinar da KDD e IHC pode trazer significantes contribuições, tal como aumentar o potencial das ferramentas utilizando-se do conhecimento do usuário e ajudando os usuários a obter resultados cada vez mais expressivos das bases de dados através das ferramentas de mineração. Entretanto, as ferramentas analisadas mostram que outros aspectos de IHC necessitam ser abordados mais cuidadosamente.

Como trabalhos futuros tem-se a ampliação do percurso cognitivo que o usuário terá que seguir para ao término responder um novo questionário de avaliação das ferramentas selecionadas. Para o percurso serão adicionadas novas tarefas de mineração de dados, abrangendo um método de cada uma das tarefas que constituem a mineração de dados (Classificação, Associação, Agrupamento e Regressão). Com isso, tem-se a intenção de avaliar mais profundamente as ferramentas e quão experiente o usuário tem que ser para poder usar estes métodos e ao fim propor uma ontologia que disponha todos os métodos de mineração de dados de forma que melhore a interação do usuário com a ferramenta.

Referências

CANVAS. Site Oficial da Ferramenta ORANGE CANVAS. Disponível em: http://orange.biolab.si/. Acesso em 27 de outubro de 2014.

EVERITT, B. S.; LANDAU, S.; MORVEN, L. Cluster Analysis. 4a ed. Londres: Hodder Arnold Publishers, 2001.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHRUSAMY, R. Advances in knowledge Discovery & Data Mining. California: AAAI/MIT, 1996.

KNIME. Site Oficial da Ferramenta KNIME. Disponível em: http://www.knime.org/. Acesso em 27 de outubro de 2014.

MACQUEEN, J. B. Some Methods for classification and Analysis of Multivariate Observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297. 1967.

RAPIDMINER. Site Oficial da Ferramenta RAPIDMINER STUDIO. Disponível em: http://rapidminer.com/products/rapidminer-studio/. Acesso em 27 de outubro de 2014.

WEKA. Waikato Environment for Knowledge Analysis. Disponível em: http://www.cs.waikato.ac.nz/ml/weka/. Acesso em 27 de outubro de 2014.

WHARTON, C., RIEMAN, J., LEWIS, C., and POISON, P. The cognitive walkthrough method: A practitioner's guide.1994.

WONG, P. C. Visual Data Mining. IEEE Computer Graphics and Applications, Los Alamitos, v.19, no. 5, p. 20-21, 1999.

Organização: Parceria: Promoção: Apoio:

ISBN: 978-85-7669-294-2 (online)