Goiânia - GO

XI Brazilian Symposium on Information Systems

Information Systems:
A Computer Socio-Technical Perspective

May 26-29, 2015
Goiânia - Goiás - Brazil

acm In-Cooperation
acm SIGMIS
acm SIGAPP
ASSOCIATION FOR INFORMATION SYSTEMS

XI Simpósio Brasileiro de Sistemas de Informação

Sistemas de Informação:
A Visão Sociotécnica da Computação

26 a 29 de Maio de 2015
Goiânia - GO - Brasil

ANAIS

VIII Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI)

Vol. 3

Publicado por:
Instituto de Informática (INF)
Universidade Federal de Goiás (UFG)
http://www.inf.ufg.br/sbsi2015

Créditos:
Capa: Ascom/UFG
Diagramação: Luciana de Oliveira Berreta, Sérgio Teixeira de Carvalho, Gustavo Teodoro Laureano

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

S612a Simpósio Brasileiro de Sistemas de Informação (20. : 2015 : Goiânia, GO)
Anais [do] XX Simpósio Brasileiro de Sistemas de Informação [recurso eletrônico] / coordenação do Comitê de Programa: Sean Wolfgand Matsui Siqueira, Sérgio Teixeira de Carvalho ; coordenação geral: Vinícius Sebba Patto, Valdemar Vicente Graciano Neto ; realização: UFG/Instituto de Informática ; promoção: Sociedade Brasileira de Computação. – Goiânia : UFG, Instituto de Informática, 2015.

Tema central: Sistemas de informação: uma visão sociotécnica da computação.
Conteúdo: v.1. Trilhas técnicas. – v.2. Workshop de iniciação científica em sistemas de informação (II WICSI). – v.3. Workshop de teses e dissertações em sistemas de informação (VIII WTDSI). – v.4. Minicursos.
Disponível em: <http://www.inf.ufg.br/sbsi2015/anais/>

1. Sistemas de recuperação da informação – Congressos. 2. Tecnologia – Serviços de informação – Congressos. 3. Internet na administração pública – Congressos. I. Siqueira, Sean Wolfgand Matsui . II. Carvalho, Sérgio Teixeira de. III. Patto,Vinícius Sebba. IV. Graciano Neto, Valdemar Vicente. V. Universidade Federal de Goiás. Instituto de Informática. VI. Sociedade Brasileira de Computação. VII. Título. VIII. Título: Sistemas de informação: uma visão sociotécnica da computação.

CDU: 004.65

II WICSI

II Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI)
Evento integrante do XI Simpósio Brasileiro de Sistemas de Informação (SBSI)
26 a 29 de Maio de 2015
Goiânia, Goiás, Brasil.

Comitês

Coordenação Geral do SBSI 2015:
Vinícius Sebba Pato (UFG)
Valdemar Vicente Graciano Neto (UFG)

Coordenação do Comitê de Programa do WTDSI 2015:
Jose Maria Nazar David (UFJF)
Luciana de Oliveira Berretta (UFG)

Comitê do Programa Científico:
Andréa M. Magdaleno (UNIRIO)
Bruno Zarpelão (UEL)
Célia G. Ralha (UnB)
Cláudia Cappelli (UNIRIO)
Clodis Boscarioli (Unioeste)
Daniel Kaster (UEL)
Fernanda Campos (UFJF)
Marco Antônio Araújo (UFJF)
Regina Braga (UFJF)
Renata Araújo (UNIRIO)
Renato Bulcão Neto (UFG)
Rita Suzana Maciel (UFBA)
Sean Siqueira (UNIRIO)
Vaninha Vieira (UFBA)

Realização
INF/UFG – Instituto de Informática/ Universidade Federal de Goiás

Promoção
Sociedade Brasileira de Computação (SBC)

Apoio
SECTEC-GO – Secretaria de Ciência, Tecnologia e Inovação do Estado de Goiás

Patrocínio Institucional
FAPEG – Fundação de Amparo à Pesquisa do Estado de Goiás
CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

Apresentação

O Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI) é um fórum dedicado à apresentação e discussão de trabalhos de mestrado e de doutorado em Sistemas de Informação (SI) desenvolvidos nos programas de pós-graduação no Brasil. O seu objetivo é propiciar um ambiente construtivo para discussões, em que os alunos possam receber uma avaliação dos seus trabalhos por pesquisadores experientes em SI e ter acesso a um panorama representativo da pesquisa em SI no país.

Nossos principais objetivos neste fórum são: (i) estimular a integração e cooperação de pesquisadores em SI; (ii) dar uma maior visibilidade às pesquisas em andamento, tanto para a comunidade acadêmica quanto para institutos de pesquisa que vêm se estabelecendo no país; e (iii) estimular a identificação de oportunidades de aplicação das propostas apresentadas nas organizações.

É com muita satisfação que apresentamos a oitava edição do WTDSI 2015, em conjunto com XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015).

Nesta edição, tivemos um total de 19 submissões. Destas, foram selecionadas para apresentação 11 propostas de dissertação de mestrado. Cada proposta foi avaliada por, no mínimo, dois revisores. As propostas selecionadas apresentam um panorama da pesquisa nos programas de pós-graduação em SI no Brasil, envolvendo uma gama de assuntos atuais e relevantes, como: SI Inteligentes, Metodologias e abordagens para Engenharia de SI, Gestão de Processos de Negócio, Aspectos Sociais e Humanos em SI, Lógica e ontologias para SI, Desenvolvimento Dirigido a Modelos (MDD), Modelagem conceitual de SI, Tecnologia da Informação no Governo Federal, Estratégia de SI e modelos inovadores de negócios.

Agradecemos aos autores dos trabalhos e seus orientadores, por prestigiarem o WTDSI 2015; aos membros do comitê de programa, pelo tempo dedicado e valiosas contribuições sugeridas aos autores; e à organização geral do SBSI 2015, por todo o suporte oferecido. Desejamos um ótimo WTDSI 2015 a todos, com ótimas discussões. Aproveitem!

Goiânia, 26 de maio de 2015.

José Maria N. David (UFJF) e Luciana de Oliveira Berretta (UFG)

Coordenadores do WTDSI 2015.

Biografia dos Coordenadores do Comitê de Programa do SBSI 2015

Jose Maria Nazar David possui graduação em Engenharia Elétrica pelo Instituto Militar de Engenharia (1983), mestrado (1991) e doutorado (2004) em Engenharia de Sistemas e Computação pela COPPE/Universidade Federal do Rio de Janeiro. Atualmente é professor adjunto e membro do Programa de Pós-graduação em Ciência da Computação da Universidade Federal de Juiz de Fora. É membro da Comissão Especial de Sistemas de Informação (CE-SI) e da Comissão Especial de Sistemas Colaborativos (CESC) da Sociedade Brasileira de Computação (SBC). Tem experiência na área de Ciência da Computação, com ênfase em Engenharia de Software, atuando principalmente nos seguintes temas: sistemas colaborativos, desenvolvimento distribuído de software, arquitetura de software, manutenção e visualização de software, informática e educação, e aprendizagem cooperativa apoiada por computador. Mais informações podem ser obtidas em http://lattes.cnpq.br/3640497501056163.

Luciana de Oliveira Berretta é graduada em Ciência da Computação pelas Faculdades Objetivo (2000). Especialista em Orientação a Objetos e Internet pela Faculdade Anhanguera (2003). Mestre em Engenharia Elétrica, pela Universidade Federal de Goiás, na área de Computação Aplicada (2005). Doutoranda em Engenharia Elétrica na Universidade Federal de Uberlândia, na área de Computação Gráfica (Realidade Virtual). Atualmente é Professora Adjunta do Instituto de Informática da Universidade Federal de Goiás, atuando principalmente nos seguintes temas: Algoritmos, Programação de Computadores e Computação Gráfica. Mais informações podem ser obtidas em http://lattes.cnpq.br/0987947348533817.

VIII Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI)

Análise de Rede de Colaboração Científica como Ferramenta na Gestão de Programas de Pós-graduação

Alternative Title: Analysis of Scientific Collaboration Network as a Management Tool for Graduate Programs

Aurélio Ribeiro Costa
Departamento de Ciência da Computação
Instituto de Ciencias Exatas
Universidade de Brasília
Brasília - DF - Brasil
arcosta@gmail.com

Célia Ghedini Ralha (Orientadora)
Departamento de Ciência da Computação
Instituto de Ciencias Exatas
Universidade de Brasília
Caixa Postal 4466 - Brasília - DF
CEP 70.904-970
ghedini@cic.unb.br

RESUMO

O desempenho de um programa de pós-graduação é aferido pela CAPES, em parte pelo seu nível de publicação, para tanto, faz-se necessário que a gestão do programa seja feita de forma a maximizar a qualidade das publicações que são realizadas pelos pesquisadores associados. No contexto dos relacionamentos de coautoria em publicações, a análise de rede de colaboração científica se mostra uma ferramenta adequada para avaliar as parcerias já formadas, bem como para estimular a formação de novas parcerias. Neste artigo é apresentada uma ferramenta de análise de rede aplicada à gestão de um programa de pós-graduação de uma Universidade Federal. Ressaltamos a utilidade ao gestor do programa, bem como aos demais pesquisadores, do módulo de recomendação de parcerias. O desenvolvimento da pesquisa usa a metodologia *Design Science Research* para guiar tanto a construção do artefato quanto para elaborar a documentação associada. Preliminarmente, utilizando dados de um programa de pós-graduação, pode-se perceber o potencial de recomendação de novas parcerias de co-autoria na rede de colaboração previamente formada.

Palavras-Chave

Sistemas de Recomendação, NoSQL

ABSTRACT

The performance of a graduate program is assessed by CAPES partly by their level of publication. Therefore, it is necessary that the program chair has instruments to analyze the quality of publication produced by the associated researchers. In the context of co-authoring relationships in publications, network analysis has been proved to be an appropriate tool to evaluate the relationships already formed, and to stimulate the formation of new relationships. This paper presents a network analysis tool applied to real data of a graduate program from a Brazilian Federal University. The data was modeled in a NoSQL graph oriented database including all associated researchers' publication. We emphasize the usefulness of the partners' recomendation module to the program chair, as well as associated researchers. The development of this research work used the *Design Science Research* methodology to guide both the construction of the artefact and the associated documentation. Preliminary, using data from a graduate program, we can note the recomendation potential to integrate new partners to the scientific network already formed.

SBSI 2015, May 26th-29th, 2015, Goiânia, Goiás, Brazil

Categorias e Descritores do Assunto

H.2.8 [**Database Applications**]: Scientific databases; H.4.2 [**Types of Systems**]: Decision support; J.1 [**ADMINISTRATIVE DATA PROCESSING**]: Education

Termos Gerais

Social Network Analysis, Design Science Research

Keywords

Recommending Systems, NoSQL

1. INTRODUÇÃO

Um dos fundamentos da gestão de um programa de pós-graduação baseia-se na maximização dos indicadores de produção acadêmica, especialmente a publicações em periódicos bem qualificados com abrangência internacional. Os indicadores de produção científica e acadêmica são elementos fundamentais nas avaliações realizadas pela CAPES[1], fundação vinculada ao Ministério da Educação responsável por avaliar os programas de pós-graduação em todos os estados da Federação. Nesse sentido, faz-se necessário que os gestores desses programas disponham de uma ferramenta que os permita visualizar a distribuição das publicações que foram realizadas pelos pesquisadores. O objetivo de tal ferramenta

[1]http://www.capes.org.br

é auxiliar no direcionamento de ações objetivando a melhoria dos indicadores de produção científica e acadêmica.

Este trabalho objetiva apresentar uma ferramenta de auxílio à gestão dos indicadores de produção acadêmica de um programa de pós-graduação. A abordagem utilizada para construção dessa ferramenta foi baseada na análise da rede de colaboração científica formada pelos pesquisadores vinculados ao programa. Também é objetivo deste trabalho detalhar a implementação da funcionalidade de recomendação de parcerias nas futuras publicações a serem desenvolvidas pelos pesquisadores vinculados ao programa.

2. APRESENTAÇÃO DO PROBLEMA

O desempenho de um programa de pós-graduação é medido, em parte, pelo nível de publicações científicas realizadas pelos pesquisadores vinculados a esse programa. Segundo o documento de área de ciência da computação de 2013 publicador pela CAPES [1], quanto à produção bibliográfica, os docentes devem estar publicando regularmente em veículos internacionais classificados nos estrato superiores (A1-B1) da plataforma Qualis CC, como requisito e orientação para proposta de novos cursos de doutorado. Assim, é imprescindível a busca pela maximização da qualidade nas publicações científicas e acadêmicas realizadas pelos pesquisadores vinculados a um programa de pós-graduação.

Uma conjectura adotada neste trabalho é fundamentada na hipótese de que uma "boa" rede de colaboração científica permite ao pesquisador melhorar seu nível de publicações acadêmicas.

3. PROPOSTA DE SOLUÇÃO

Para o desenvolvimento da pesquisa está sendo usado a metodologia *Design Science Research* defendida por [2], a qual envolve a geração de conhecimento novo através do desenvolvimento de artefatos inovadores e da análise do uso e/ou do desempenho de tais artefatos por meio de reflexão e abstração. Tais artefatos incluem, mas não se limitam a, interfaces homem/maquina, algoritmos, metodologias de projeto de sistemas e linguagens.

A solução proposta para o problema apresentado na Seção 2 consiste na construção de um modelo de rede de colaboração científica através de um grafo, e a implementação de um artefato para análise dessa rede bem como para sugestão de novos relacionamentos entre pesquisadores de tal maneira que permita a melhoria dos níveis de produção científica desses pesquisadores.

O modelo de rede de colaboração está sendo persistido em um banco de dados NoSQL orientado a grafo.

4. PROJETO DE AVALIAÇÃO DA SOLUÇÃO

A avaliação da solução proposta será realizada através da confecção de um questionário a ser disponibilizado aos pesquisadores cujos currículos foram processados pela ferramenta para que possa ser realizada uma avaliação das recomendações de parcerias realizadas.

5. ATIVIDADES JÁ REALIZADAS

A pesquisa iniciou-se com uma revisão bibliográfica acerca do problema de maximização da qualidade das publicações destacando-se o realizado por [5], o qual usa como fonte de informação a plataforma Lattes. Foram pesquisados ainda temas pertinentes ao desenvolvimento da pesquisa como bancos de dados NoSQL orientados a grafo[6], sistemas de recomendação[7] bem como *Design Science Research*[8].

Para a construção do grafo de relacionamento, construto inicial do *Design Science*, foram adotados dois tipos de nós, o nó pesquisador e o nó publicação. Também foi definido um tipo de aresta, o qual representa a relação de autoria de uma publicação. O ponto de partida para compreender a relação de parceria entre pesquisadores foi realizado através da coleta de dados dos currículos Lattes dos professores vinculados ao Departamento de Ciência da Computação da Universidade de Brasília e atuantes no programa de pós-graduação em Informática, os quais se relacionam através de autoria em artigos completos publicados em periódicos.

As relações de colaboração foram extraídas através dos autores dos artigos cadastrados na plataforma Lattes. Foi utilizado como chave de acesso no banco de dados os nomes dos autores dos artigo e os nomes utilizados para referência de cada autor na Seção Identificação do currículo Lattes. Assim, com os nomes dos autores dos artigo e os nomes utilizados para referência de cada autor, foi possível a geração das arestas, representando as autorias do grafo permitindo então a visualização das parcerias já realizadas.

A construção do artefato para análise da rede de colaboração científica foi iniciada com a modelagem das informações contidas nos currículos Lattes dos pesquisadores, mais especificamente dos dados referentes às publicações em periódicos. Em seguida, foi definido o fluxo de tratamento desses dados desde a extração na plataforma Lattes até a visualização da informação conforme Figura 1.

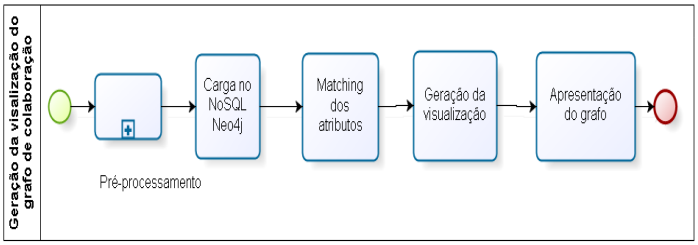

Figura 1: Processo de carga e visualização

Após o sub-processo de pre-processamento, detalhado na Figura 2, a tarefa de carga no Neo4j é iniciada, quando será realizado o *matching* dos atributos de autoria para criação da rede de colaboração. Na sequencia, ocorrem a geração da visualização, quando são construídos os gráficos. Posteriormente é apresentado o grafo para o usuário.

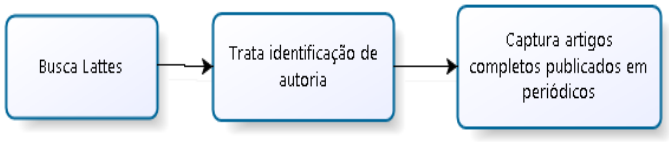

Figura 2: Sub processo de pré processamento

6. CONCLUSÃO

Preliminarmente, utilizando dados do Programa de Pós-graduação em Informática da Universidade de Brasília, pode-se perceber o potencial de recomendação de novas parcerias de co-autoria na rede de colaboração previamente formada.

Dando continuidade a pesquisa, pretende-se validar as recomendações realizadas pelo protótipo do módulo de recomendação já implementado, assim como expandir a base de dados utilizada de forma a abranger pesquisadores de diversos programas nacionais e internacionais. Pretende-se ainda experimentar outras abordagens no modelo de recomendação, principalmente a adoção de sistemas multiagentes em um modelo multi-camadas, no qual cada camada representa uma forma de interação entre pesquisadores como coautoria, participação em projeto, participação em banca apenas para citar algumas possibilidades.

7. REFERÊNCIAS

[1] Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Documento de Área 2013 - ciência da computação. `http://www.capes.gov.br/component/content/article?id=4656:ciencia-da-computacao`, Oct 2013.

[2] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in information systems research. *MIS Q.*, 28(1):75–105, March 2004.

[3] Victor Ströele, Geraldo Zimbrão, and Jano M. Souza. Group and link analysis of multi-relational scientific social networks. *J. Syst. Softw.*, 86(7):1819–1830, July 2013.

[4] Jim Webber. A programmatic introduction to neo4j. In *Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity*, SPLASH '12, pages 217–218, New York, NY, USA, 2012. ACM.

[5] Renato Balancieri, Alessandro Botelho Bovo, Vinícius Medina Kern, Roberto Carlos dos Santos Pacheco, and Ricardo Miranda Barcia. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na plataforma lattes. *Revista IBICT*, 34(1), 2005.

[6] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O'Reilly, 2013.

[7] Prem Melville and Vikas Sindhwani. Recommender systems. In *Encyclopedia of machine learning*, pages 829–838. Springer, 2010.

[8] Vijay K Vaishnavi, William Kuechler, and William Kuechler Jr. *Design science research methods and patterns: innovating information and communication technology*. Crc Press, Oct, 30 2007.

[9] Claudio Tesoriero. *Getting Started with OrientDB*. Packt Publishing Ltd, 2013.

[10] Matías Javier Antiñanco. *Bases de Datos NoSQL: escalabilidad y alta disponibilidad a través de patrones de diseño*. PhD thesis, Facultad de Informática, Universidad Nacional De La Plata, 2014.

[11] Dan Brickley and Ramanathan V Guha. Resource description framework (rdf) schema specification 1.0: W3c candidate recommendation 27. March 2000.

Aplicação da Ciência dos Dados em Hidrologia para Estimativa da Evapotranspiração

Alternate Title: Application of Data Science in Evapotranspiration Estimation

Fernando Xavier
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Avenida Pasteur, 458 - Urca
Rio de Janeiro – RJ – Brasil
(+55) 21-96672-0377
fernando.xavier@uniriotec.br

Asterio Kiyoshi Tanaka
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Avenida Pasteur, 458 - Urca
Rio de Janeiro – RJ – Brasil
(+55) 21-2541-4959
tanaka@uniriotec.br

RESUMO

Estudos relacionados aos recursos hídricos têm grande importância em muitas atividades, como irrigação, abastecimento de água e geração de energia. Estes estudos podem gerar um grande e complexo volume de dados que, muitas vezes, são difíceis de analisar. Este trabalho visa aplicar técnicas de Ciência dos Dados na análise desses dados, com o objetivo de resolver um problema no campo da Hidrologia: a estimativa da evapotranspiração. Espera-se que este trabalho contribua para pesquisas em Hidrologia, ao propor uma nova forma de analisar estes dados, e em Sistemas de Informação, demonstrando como a Ciência dos Dados pode ser aplicada para resolver problemas de diferentes áreas de pesquisa.

Palavras chaves

Evapotranspiração, Mineração de Dados, Hidrologia, Ciência dos Dados.

ABSTRACT

Studies related to water resources have importance in many activities, such as irrigation, water supply and power generation. These studies can generate a large and complex amount of data that are often difficult to analyze. This work aims at applying Data Science techniques for analysing such data, to solve a problem in the Hydrology field: the evapotranspiration estimation. It is hoped that this work will contribute to researches in Hydrology, by proposing a new way to analyze these data, and in Information Systems, by demonstrating how Data Science can be applied for solving problems from the different research domains.

Categories and Subject Descriptors

H.2.8 [**Information Systems**]: Database Applications – *data mining*.

General Terms

Algorithms, Experimentation, Theory.

Keywords

Evapotranspiration, Data Mining, Hydrology, Data Science.

1. INTRODUÇÃO

A água é um recurso fundamental para o planeta e existem diversas iniciativas relacionadas ao tema, com diversos interesses como: desastres naturais, agricultura, abastecimento, entre outros. A Organização das Nações Unidas, por exemplo, mantém o programa *Intergovernamental Panel on Climate Change* (IPCC), que busca analisar a produção científica realizada no mundo relativa às mudanças climáticas [9]. Além do IPCC, existem diversos órgãos que mantém dados meteorológicos e climáticos, utilizados por pesquisadores para compreensão dos fenômenos naturais, permitindo a análise de cenários futuros através da geração de modelos preditivos, que também geram dados.

A quantidade de dados históricos e os gerados pelos modelos preditivos tende a ser cada vez maior e mais complexa, conforme estudo realizado por Overpeck et al em [15]. Este estudo indica uma previsão de que os dados climáticos (reais e simulados) deverão superar a casa dos 350 *Petabytes* em 2030.

Essa massa de dados está relacionada ao que se chama de era do *Big Data* [3], que se refere ao volume de dados que devem ser adquiridos e processados, a complexidade dos dados e a taxa de geração desses [15]. Nesse cenário, o desafio para os pesquisadores é como analisar essa massa de dados com velocidade para prover tempo útil e valor aos resultados.

Nesse contexto se insere a Ciência dos Dados que, com o uso de disciplinas de Computação, Estatística e Matemática, poderá diminuir a dificuldade de manipulação e visualização dessa massa de dados [13] [17]. Uma característica da Ciência dos Dados, como pesquisa em Sistemas de Informação, é a sua natureza transdisciplinar, e a era do *Big Data* representa uma oportunidade para os pesquisadores de Ciência dos Dados [2].

Para demonstrar a importância e os ganhos que o uso da Ciência dos Dados pode trazer a diversas áreas, propõe-se neste trabalho de pesquisa a aplicação da Ciência dos Dados a um problema da

área da Hidrologia: a estimativa da evapotranspiração, um importante componente do ciclo hidrológico.

Os modelos matemáticos existentes para essa estimativa dependem de dados que nem sempre estão disponíveis ou são simples de obter [12]. Nesse sentido, outros métodos são propostos, como o uso de redes neurais e sensoriamento remoto que, no entanto, também têm limitações, como baixa acurácia nos resultados e a fixação de quais dados devem estar disponíveis.

Com base nessas limitações, tem-se como objetivo de pesquisa simplificar o processo de estimativa da evapotranspiração. Com a aplicação da Ciência dos Dados, espera-se gerar modelos de estimativa da evapotranspiração independentemente de quais dados estejam disponíveis em uma região. Essa hipótese será verificada com a comparação dos valores calculados pelos modelos gerados com os valores históricos, quando disponíveis.

Na próxima seção, é feita a apresentação do problema, seguida pela proposta de solução na Seção 3. Na Seção 4, é apresentado como a solução proposta será avaliada. Na Seção 5, serão descritas as atividades realizadas ou em andamento nesta pesquisa e, por fim, na Seção 6 são feitas as considerações finais, indicando as contribuições esperadas, limites e possíveis desdobramentos.

2. APRESENTAÇÃO DO PROBLEMA

A evapotranspiração é definida como a soma da evaporação da água de rios ou lagos e a transpiração da vegetação, que retorna para a atmosfera na forma de vapor [4]. A Figura 1 ilustra a evapotranspiração e sua relação no ciclo hidrológico.

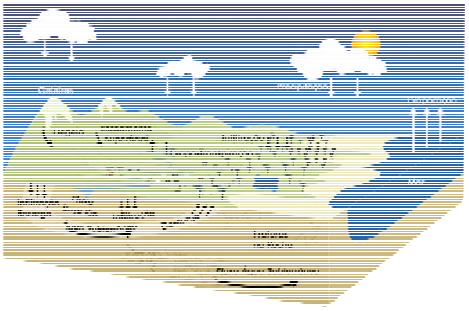

Figura 1. Ciclo Hidrológico [14]

A água, que chega aos continentes através de precipitações, pode seguir alguns caminhos: infiltrar no solo ou nas rochas, formar aquíferos, ressurgir na superfície ou alimentar lagos e rios. Nos casos em que a capacidade de absorção do solo é menor do que a taxa de precipitação, a água pode escoar pela superfície. A água da superfície, dos rios e lagos retorna à atmosfera através da evaporação, em conjunto com a água liberada pelas plantas através da transpiração, processo conhecido como evapotranspiração [14]. De volta à atmosfera, formam-se nuvens que, quando carregadas, geram precipitações, reiniciando o ciclo.

O valor da evapotranspiração, medido em milímetros/dia, é fundamental no planejamento de atividades como a irrigação, que representam mais de 75% do uso mundial de recursos hídricos [11]. A quantidade da água a ser irrigada depende da quantidade necessária para as plantações, que varia de acordo com a água que retornou à atmosfera pela evapotranspiração.

Devido ao fato da medida direta da evapotranspiração ser difícil e onerosa, a abordagem corrente é sua estimativa através de métodos matemáticos [4]. A equação de Penman-Monteith (PM) é o método de referência da Organização das Nações Unidas para Alimentação e Agricultura (FAO) [6] e usa dados climáticos para estimar a evapotranspiração, que nem sempre são fáceis de obter ou não estão disponíveis [12].

Pelo fato da equação de PM não poder ser utilizada na indisponibilidade do valor de alguma variável, algumas alternativas foram propostas, como o uso de redes neurais [10] e sensoriamento remoto [12]. As pesquisas usando redes neurais, apesar de simplificarem o processo, têm utilizado variáveis definidas. Já os métodos de sensoriamento remoto trazem como vantagem a estimativa da evapotranspiração para áreas maiores, embora tenham algumas limitações, como ajustes para regiões montanhosas, valores dependendo da resolução dos sensores e a acurácia da evapotranspiração estimada estar entre 10-15% em relação a medidas *in situ* [12].

Nesse cenário, define-se como problema de pesquisa a seguinte questão: É possível simplificar a estimativa da evapotranspiração independentemente de quais variáveis tenham valores? A hipótese decorrente dessa questão é que a evapotranspiração poderá ser estimada, mesmo que o conjunto de dados (*dataset*) utilizado não tenha todos os valores disponíveis.

3. PROPOSTA DE SOLUÇÃO

Com o objetivo de simplificar as abordagens existentes na estimativa da evapotranspiração, propõe-se o uso de técnicas da Ciência dos Dados, tendo os dados como fonte para aprendizado dos padrões e definição de modelos, que podem ser usados para geração de novos dados [3].

Através de mineração dos dados meteorológicos, provenientes de estações de medição, serão gerados modelos de estimativa da evapotranspiração para cada uma dessas estações. Ou seja, o modelo será gerado de acordo com os dados existentes da estação.

Os dados a serem utilizados estão disponíveis no banco de dados do Instituto Nacional de Meteorologia (INMET) [8], que contêm séries históricas de dados meteorológicos de 265 estações de medição no Brasil, com valores de diversas variáveis, como temperatura, umidade, pressão, dentre outras. Além disso, as séries históricas do INMET contêm, desde 2006, valores estimados da evapotranspiração, que serão usados para avaliação da proposta de solução. Nem todas as estações possuem os *datasets* completos, o que será utilizado para avaliar o comportamento da abordagem na ausência de dados.

Após a mineração dos dados disponíveis originalmente, serão feitas manipulações nos *datasets* para se retirar variáveis, na atividade chamada de Pré-Processamento do processo de Descoberta de Conhecimento em Banco de Dados [1] [7]. O objetivo da manipulação do *dataset* é verificar o comportamento da abordagem em novo processamento da mineração de dados, para uma mesma estação na ausência dos valores de uma variável.

4. AVALIAÇÃO DA SOLUÇÃO

Os resultados gerados por este trabalho serão comparados com os valores estimados pela equação de PM, quando for possível sua utilização, e dados de medição direta, quando estiverem disponíveis. Os valores gerados também serão comparados com os valores presentes nas séries históricas do INMET.

Além disso, a variedade da disponibilidade dos valores das variáveis entre as estações também será usada para avaliar a

solução proposta, analisando-se a precisão da evapotranspiração estimada em relação a outras estações, cujas variáveis com valores disponíveis sejam diferentes. Caso a precisão entre o valor estimado pela abordagem proposta e o valor estimado na série histórica do INMET tenha pouca variação, mesmo com variáveis sem valores preenchidos, então a solução terá comprovada sua eficácia, ao garantir que a estimativa da evapotranspiração poderá ser feita com os valores que estiverem disponíveis. A hipótese será falsa se os valores da evapotranspiração só puderem ser estimados com boa acurácia se o *dataset* estiver completo.

Será avaliada, também, a precisão da estimativa de acordo com as variáveis com valores disponíveis nas séries históricas do INMET. Essa avaliação poderá indicar quais variáveis têm maior relevância na estimativa da evapotranspiração.

5. ATIVIDADES JÁ REALIZADAS

- Revisão de Literatura: está sendo feito um mapeamento sistemático da literatura, conforme os passos descritos por Santos [16], para conhecer as abordagens utilizadas;
- Experimento: Usando dados meteorológicos das estações do INMET no Estado do Rio de Janeiro, foi feito um experimento usando mineração de dados para geração de modelos de estimativa para a evapotranspiração, com resultados promissores [18].

6. CONSIDERAÇÕES FINAIS

A abordagem proposta neste trabalho é a principal contribuição, com a aplicação de técnicas de Ciência dos Dados em um problema da área da Hidrologia e Agricultura.

O processo de estimativa de evapotranspiração tem importância fundamental na irrigação, já que a má gestão dos recursos hídricos pode gerar problemas como desperdício de água que, segundo a FAO, poderá afetar até dois terços da população mundial em 2050 [5]. Espera-se que o método proposto neste trabalho possa auxiliar agricultores na melhor gestão do uso dos recursos hídricos nas atividades de irrigação.

O protocolo gerado no mapeamento sistemático da literatura também será uma contribuição, pois possibilitará novos mapeamentos sistemáticos, permitindo atualizar e complementar os resultados já identificados [16].

Espera-se também que a abordagem proposta neste trabalho seja útil para o planejamento de instalação de novas estações de medição, com a aquisição apenas dos instrumentos de medição das variáveis que tenham mais peso para estimativa da evapotranspiração, de acordo com a precisão desejada.

Além disso, pode-se verificar se os modelos gerados neste trabalho podem ser utilizados em áreas não monitoradas, cujas características físicas sejam semelhantes às de áreas monitoradas. Uma sugestão seria a utilização de técnicas de agrupamento por similaridade, usando as características físicas de cada região.

7. REFERÊNCIAS

[1] Affonso, M. A., Revoredo, K., & Andrade, L. (2012). Avaliando uma Oportunidade Exploratória de Petróleo através de Mineração de Dados. VIII Simpósio Brasileiro de Sistemas de Informação (SBSI 2012)

[2] Agarwal, R., Vasant, D. (2014) Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research 25(3):443448.

[3] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, 36(4), 11651188.

[4] Di Bello, R. C. (2005). Análise do Comportamento da Umidade do Solo no Modelo ChuvaVazão Smap II–Versão com Suavização Hiperbólica Estudo de Caso: Região de Barreiras na Bacia do Rio Grande-BA (Dissertação de Mestrado, Universidade Federal do Rio de Janeiro).

[5] EBC (2015) FAO: falta de água afetará dois terços da população mundial em 2050. Disponível em http://agenciabrasil.ebc.com.br/internacional/noticia/2015-04/fao-falta-de-agua-afetara-dois-tercos-da-populacao-mundial-em-2050. (Acessado em abril/2015).

[6] FAO (2015). Organização das Nações Unidas para Alimentação e Agricultura. Chapter 2 - FAO Penman-Monteith Equation (Acessado em janeiro/2015)

[7] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37

[8] INMET (2014). Instituto Nacional de Meteorologia. Disponível em http://www.inmet.gov.br (acessado em Novembro/2014).

[9] IPCC (2014), Intergovernmental Panel on Climate Change http://www.ipcc.ch/organization/organization.shtml (Acessado em novembro/2014).

[10] Kumar, M., Raghuwanshi, N. S., & Singh, R. (2011). Artificial neural networks approach in evapotranspiration modeling: a review. Irrigation science, 29(1), 11-25.

[11] Kumar, R., Jat, M. K., & Shankar, V. (2012). Methods to estimate irrigated reference crop evapotranspiration a review. Water Science and Technology, 66(3), 525.

[12] Liou, Y. A., & Kar, S. K. (2014). Evapotranspiration estimation with remote sensing and various surface energy balance algorithms—A review. Energies, 7(5), 2821-2849.

[13] Mattmann, C. A. (2013). Computing: A vision for data science, Nature, 493, 473–475

[14] MMA (2015). Ciclo Hidrológico, Ministério do Meio Ambiente. Disponível em http://www.mma.gov.br/agua/recursos-hidricos/aguas-subterraneas/ciclo-hidrologico (Acessado em março/2015)

[15] Overpeck, J. T., Meehl, G. A., Bony, S., & Easterling, D. R. (2011). Climate data challenges in the 21 st century. Science(Washington), 331(6018), 700702.

[16] Santos, G., (2008), Ambientes de Engenharia de Software Orientados à Corporação. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

[17] Vasant, D. (2013), Data Science and Prediction. Communications of the ACM, vol. 56, no. 12

[18] Xavier, F.; Tanaka. A.K.; Revoredo, K.C. Aplicação de Descoberta de Conhecimento em Bases de Dados na Estimativa da Evapotranspiração: um Experimento no Estado do Rio de Janeiro. XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015). Aceito para publicação.

Aprendizado automático de ontologias a partir de Data Warehouses

Alternative title: Automatic ontology learning from Data Warehouses

Tiago Outerelo da Silva
PPGI UNIRIO
Avenida Pasteur 458 – Urca –
Rio de Janeiro - RJ
tiago.dasilva@uniriotec.br

Fernanda Baião
PPGI UNIRIO
Avenida Pasteur 458 – Urca –
Rio de Janeiro - RJ
fernanda.baiao@uniriotec.br

Kate Revoredo
PPGI UNIRIO
Avenida Pasteur 458 – Urca –
Rio de Janeiro - RJ
katerevoredo@uniriotec.br

RESUMO

Business Intelligence (BI) oferece meios para fornecer informações e derivar conhecimento através de ferramentas de análise para tomada de decisão, mas comumente não apresentam uma representação formal de conhecimento que explicite e descreva semanticamente os dados armazenados no Data Warehouse (DW). Assim sendo, ontologias são artefatos que representam um domínio através dos seus conceitos e relacionamentos. Este artigo propõe a geração automática de ontologias em sistemas de BI, utilizando regras de mapeamento entre os elementos de DWs e os elementos destas ontologias. As principais contribuições desta proposta são a criação e adaptação dessas regras de mapeamento e a criação de um processo automatizado para a geração de ontologias a partir de DWs.

Palavras-chave

Data Warehouse, aprendizado de ontologia.

ABSTRACT

Business Intelligence (BI) provide the means to provide information and derive knowledge through analysis for decision-making tools, but often have no formal knowledge representation that clarifies and semantically describe the data stored in Data Warehouse (DW). Thus, ontologies are artifacts that represent a domain through its concepts and relationships. This article proposes the automatic generation of ontologies in BI systems, using mapping rules between DWs elements and the elements of these ontologies. The main contributions of this proposal are the creation and adaptation of these mapping rules and the creation of an automated process for generating ontologies from DWs.

Categories and Subject Descriptors

H.2.1 [**Database Management**]: Logical Design – *Data models*.
I.6.5 [**Simulation and Modeling**]: Model Development – *Modeling methodologies*.

General Terms

Algorithms, Management, Design

Keywords

Data Warehouse, ontology learning.

1. INTRODUÇÃO

As organizações estão sobrecarregadas com o volume crescente de dados que são continuamente gerados e armazenados em repositórios corporativos [1], que precisam ser analisados quando uma decisão precisa ser tomada, tais como estratégias de negócio para enfrentar concorrentes, estratégias para definição de preços de produtos e tendências de comportamento de clientes [2].

Soluções de Business Intelligence (BI) oferecem meios para fornecer informações e derivar conhecimento através de ferramentas de análise para tomada de decisão [3]. Elas auxiliam na análise de grandes volumes de dados, transformando-os em informação significativa, útil e esclarecedora. Podemos definir BI como um conjunto de teorias, metodologias, arquiteturas e tecnologias com o objetivo de recuperar e transformar dados brutos em informações significativas e úteis, permitindo que os níveis operacional, tático e estratégico de uma organização possam tomar decisões melhores e de forma mais ágil [5].

A modelagem multidimensional é a técnica de modelagem de dados utilizada em ambientes convencionais de BI e é orientada à análise de assuntos. Um banco de dados estruturado com essa técnica de modelagem é chamado de Data Warehouse (DW). Os elementos básicos dos modelos multidimensionais são os fatos e as dimensões. Fatos são estruturas de dados onde são armazenadas as métricas que se quer analisar e dimensões são estruturas de dados que representam as visões (assuntos) de análise sobre os fatos.

Em um DW implementado em um SGBD relacional, as tabelas de fato são relacionadas a tabelas de dimensão. Em modelos do tipo estrela, as dimensões são tabelas desnormalizadas e cada uma pode armazenar vários níveis de uma mesma análise. Um exemplo seria a dimensão Tempo, que pode armazenar numa só tabela as análises de dia, mês e ano. Em modelos do tipo floco de neve, as dimensões são tabelas normalizadas e cada uma é um nível de análise. Usando o exemplo anterior, teríamos uma tabela para a análise de dia, relacionada a outra tabela para o mês, e essa relacionada a outra para o ano. Como exemplo, imaginemos uma aplicação de BI sobre os funcionários de uma empresa. Uma representação do conhecimento sobre essa aplicação possibilitaria a um analista de negócio ou um analista de BI saber as métricas disponíveis para consulta, por quais visões de análise estão disponíveis (granularidade) e as relações entre essas visões (hierarquias).

Assim sendo, ontologias são artefatos que representam um domínio através dos seus conceitos e relacionamentos. Uma ontologia pode ser definida como uma especificação formal e explícita de uma conceituação compartilhada, sendo amplamente utilizada para uma representação formal mais rica de conhecimento para ser interpretável por máquina [14].

Segundo Guarino [13], podemos classificar ontologias como de alto nível, que descrevem conceitos mais genéricos, de domínio ou de tarefa, que especializam conceitos de alto nível e descrevem o vocabulário de um domínio ou de uma tarefa, respectivamente, ou de aplicação, que especializam conceitos de domínio e tarefa e descrevem o vocabulário de uma aplicação.

No entanto, a maioria dos sistemas de informação, atualmente, não apresentam uma ontologia que descreva suas informações de negócio, pelo fato de que sua construção não faz parte do ciclo de vida de desenvolvimento de software e do processo manual de construção de ontologias ser uma tarefa difícil, dispendiosa, demorada e requerer profundo conhecimento do domínio [7]. Sistemas de Business Intelligence também são sistemas de informação e apresentam a mesma questão de ausência de ontologias associadas a suas implementações.

Além desta seção de introdução, este artigo é composto das seguintes seções: a seção 2 apresenta o problema endereçado pela pesquisa em questão, a seção 3 apresenta a proposta de solução para o problema endereçado, a seção 4 descreve o projeto de avaliação da solução e a seção 5 relata as atividades já realizadas na pesquisa. A seção 6 contém as considerações finais.

2. APRESENTAÇÃO DO PROBLEMA

Apesar da importância para as organizações das ferramentas analíticas providas pelas soluções de BI, existem desafios para alavancar seu impacto no processo de tomada de decisão. Esses desafios incluem a dificuldade no alinhamento da aplicação aos requisitos do negócio, a dificuldade na análise e interpretação dos dados e a falta de flexibilidade para personalizar a aplicação de acordo com o usuário [3]. Argumenta-se que esses problemas ocorrem devido à falta de integração da semântica de negócios para as fundações de ferramentas analíticas, como suas regras de negócio e fontes de informação usadas [3]. Para a realização desta integração, é necessária uma representação formal que descreva semanticamente os conceitos implementados na aplicação de BI.

Entretanto, sistemas de BI comumente não apresentam uma representação formal de conhecimento que explicite e descreva semanticamente os dados e metadados armazenados no DW.

3. PROPOSTA DE SOLUÇÃO

Para endereçar o problema apresentado acima, ontologias podem ser utilizadas para descrever a semântica de dados e metadados e fazer seu significado explícito [6]. Elas forneceriam ao sistema de BI uma representação rica, explícita e atualizável do conhecimento armazenado em seu DW.

Voltando ao exemplo de uma aplicação de BI sobre os funcionários de uma empresa, sua ontologia poderia ter a representação da métrica **Salário**, associada à hierarquia **Tempo** e contida pelo fato **Funcionário**. Essas entidades seriam representadas como classes interligadas na ontologia.

Apesar de existirem na literatura diversas abordagens de aprendizado automático de ontologias [8] [9] [12], tais abordagens requerem a existência de outras fontes de dados, como modelos de dados e ontologias para alinhamento. A utilização de outra fonte de dados para o aprendizado de uma ontologia nesse contexto, que não uma estrutura de dados multidimensional, demandaria a existência de documentação que seja computável da aplicação em questão e que a mesma esteja sempre em sincronia com os conceitos implementados ao longo do ciclo de vida do sistema. Não é possível garantir que exista uma outra fonte de informação disponível para utilização na geração de uma ontologia, nem que a mesma esteja sempre atualizada. Entretanto, em uma aplicação de BI baseada num DW, a estrutura de dados multidimensional faz parte da aplicação e seria mais uma fonte a ser utilizada, de forma a complementar outras mais tradicionais.

Como vantagem, a utilização do DW para a geração de ontologia proporciona uma fonte de informação compartilhada com a aplicação de BI, garantindo o alinhamento entre os conceitos implementados na aplicação e os conceitos do domínio que a ontologia se propõe a representar. Além disso, o modelo dimensional permite a inferência de conceitos e relações próprias do domínio de BI, como agregações e hierarquias. As características dos dados armazenados, como volume e esparsidade, também podem ser utilizados para inferir conhecimento. Por exemplo, uma tabela agregada de funcionários por faixa etária tende a ter menor volume que uma tabela fato no nível de funcionário ou idade.

A geração de ontologia a partir de modelos multidimensionais apresenta como desafios algumas questões que são inerentes a características comumente encontradas em aplicações de BI, que são o grande volume de dados armazenados nas estruturas de dados, que dificultam a manipulação dos dados armazenados no repositório e de sua estrutura, e a desnormalização dos modelos de dados, que dificultam a identificação da relação entre as classes e suas propriedades.

Essa geração da ontologia deve ser automática devido aos problemas relativos à sua construção manual e para que seja facilitada a atualização da ontologia ao longo do ciclo de vida da aplicação. Um exemplo é a realização de manutenções evolutivas no sistema, incluindo novos fatos e dimensões, que implicam na atualização da documentação existente.

Para preencher esta lacuna, a proposta de solução é a criação e aperfeiçoamento de regras de mapeamento para a geração automática de ontologias de aplicação de sistemas de BI a partir de Data Warehouses. As regras de mapeamento a serem aperfeiçoadas para utilização serão, inicialmente, as regras definidas por Prat et al. [8] [9]. Tais regras serão aperfeiçoadas para abordar aspectos mais específicos de modelagem multidimensional, como dimensões de modificação lenta, por exemplo, e com a utilização de dados e metadados presentes nas estruturas de dados do DW.

A hipótese é que, através do uso de regras de mapeamento específicas, é possível gerar ontologias de aplicação a partir de DWs. As ontologias obtidas devem contemplar não somente o conhecimento já explícito nas estruturas de dados (como traduzir tabelas para classes, por exemplo), mas também a semântica que está implícita (como categorizações de classes, por exemplo).

O primeiro passo do processo de geração da ontologia será a geração de conceitos na ontologia que refletirão o modelo de dados multidimensional, a partir do esquema de banco de dados do DW. Posteriormente, novos conceitos são adicionados à ontologia através do uso das regras de mapeamento. Para isso, serão utilizados como elementos de entrada para as regras: o esquema do banco de dados, um metamodelo de tarefa OLAP, com conceitos predefinidos, e um metamodelo do domínio, composto pelos metadados do DW, dicionários de dados e padrões de nomenclatura. A utilização destas atividades em sequência funcionará como um processo para a geração da ontologia da aplicação.

4. PROJETO DE AVALIAÇÃO DA SOLUÇÃO

Para avaliação da solução proposta, serão obtidos diferentes cenários e domínios de aplicação e realizados experimentos para

validação individual de cada regra criada ou aperfeiçoada. A validação de resultado será com um especialista, que a depender da regra pode ser um analista de BI ou um analista de negócio. Para a escolha dos DWs a serem utilizados, será levada em consideração a diversidade de possibilidades que podem ocorrer numa modelagem multidimensional, com o intuito de avaliar a abrangência da solução, tomado o cuidado de que a amostra seja estatisticamente significativa e avaliada a qualidade dos dados disponíveis. Dentre essas as características que devem ser abordadas no objeto de estudo, destacamos a presença de fatos com modelagem estrela e floco de neve, agregações e hierarquias de dimensões.

5. ATIVIDADES JÁ REALIZADAS

Foi realizado levantamento à procura de trabalhos relacionados ao problema em questão, sendo encontradas algumas pesquisas que já exploraram o aprendizado de ontologias a partir de estrutura de dados. Prat et al. [8] [9] abordam a geração de ontologia OWL-DL a partir de modelo de dados multidimensional, mas não contemplam os dados e metadados do Data Warehouse, tendo como premissa a existência de um modelo de dados conceitual. Gil et al. [10] [11] apresentam uma metodologia sistêmica para aprendizado de ontologia (SMOL), composta de fases ao longo de um processo/fluxo estruturado, entretanto não são contempladas técnicas ou métodos para a geração da ontologia e as etapas do processo não são detalhadas.

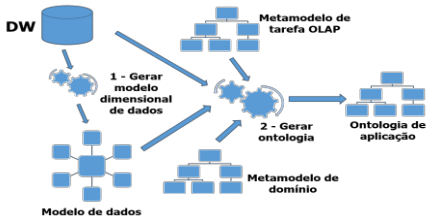

Figura 1. Processo proposto para geração de ontologia

El Idrissi et al. [7] apresentam um estudo prático de métodos que utilizam estruturas de bancos de dados como entrada para o processo de aprendizado de ontologia, porém chega à conclusão de que não existe ferramenta que extraia automaticamente uma ontologia de aplicação a partir de estrutura de banco de dados. Dou et al. [12] propõem um framewok que prevê a descoberta automática de mapeamento entre esquemas de bancos de dados e ontologias, além de um algoritmo de tradução de consultas, entretanto, a proposta não prevê a geração de ontologia, mas espera que já exista uma disponível para a realização de mapeamento com a estrutura no banco de dados. Seu foco é maior no alinhamento entre os conceitos já estabelecidos. Resumindo, não foram encontradas publicações apresentando o aprendizado de ontologias a partir de DWs de forma automática.

6. CONCLUSÃO

Neste artigo foi apresentado o objetivo de geração automática de ontologias em sistemas de BI a partir de Data Warehouses, endereçando o problema de que, comumente, esses sistemas não apresentam uma representação formal de conhecimento que explicite e descreva semanticamente os dados e metadados armazenados no banco de dados.

As contribuições desta proposta são a criação e aperfeiçoamento de regras de mapeamento entre os elementos de Data Warehouses, com seus modelos correspondentes, e os elementos de ontologias de aplicação e a implementação de uma ferramenta, com a utilização todas as regras de mapeamento, para a geração automática de ontologias (Figura 1).

7. REFERÊNCIAS

[1] Sidorova, A., and Torres, R. 2014. Business Intelligence and Analytics: A Capabilities Dynamization View. In: *Twentieth Americas Conference on Information Systems*, Georgia, 2014.

[2] Andoh-Baidoo, F., Villa, A., Aguirre, Y., and Kasper, G. 2014. Business Intelligence & Analytics Education: An Exploratory Study of Business & Non-Business School IS Program Offerings. In: *Twentieth Americas Conference on Information Systems*, Savannah, Georgia, 2014.

[3] Sell, D. et al. 2011. Adding Semantics to Business Intelligence: Towards a Smarter Generation of Analytical Tools. In: *BUSINESS INTELLIGENCE–SOLUTION FOR BUSINESS DEVELOPMENT*, p. 33, 2011.

[4] Anjariny, A. H., Zeki, A. M., and Hussin, H. 2012. Assessing Organizations Readiness toward Business Intelligence Systems: A Proposed Hypothesized Model. In: *Advanced Computer Science Applications and Technologies (ACSAT)*, 2012 International Conference on (pp. 213-218), IEEE.

[5] Airinei, D., and Homocianu, D. 2009. DSS vs. business intelligence. In: *Revista Economica*.

[6] Tong, G. et al. 2009. Application of Ontology-Based Information Integration on BI System. In: *Software Engineering*, 2009. WCSE'09. WRI World Congress on. IEEE, 2009. p. 171-175.

[7] El Idrissi, B., Baina, S. and, Baïna, K. 2013. Automatic generation of ontology from data models: a practical evaluation of existing approaches. In: *Research Challenges in Information Science (RCIS)*, 2013 IEEE Seventh International Conference on (pp. 1-12). IEEE.

[8] Prat, N., Akoka, J., and Comyn-Wattiau, I. 2012. Transforming multidimensional models into OWL-DL ontologies. In: *Research Challenges in Information Science (RCIS)*, 2012 Sixth International Conference on (pp. 1-12).

[9] Prat, N., Megdiche, I., and Akoka, J. 2012. Multidimensional models meet the semantic web: defining and reasoning on OWL-DL ontologies for OLAP. In: *Proceedings of the fifteenth international workshop on Data warehousing and OLAP* (pp. 17-24). ACM.

[10] Gil, R., and Martin-Bautista, M. J. 2014. SMOL: a systemic methodology for ontology learning from heterogeneous sources. In: *Journal of Intelligent Information Systems*, 42(3), 415-455.

[11] Gil, R., Martín-Bautista, M. J., and Contreras, L. 2010. Applying an ontology learning methodology to a relational database: University case study. In: *Semantic Computing (ICSC)*, 2010 IEEE Fourth International Conference on (pp. 313-316). IEEE.

[12] Dou, D., Qin, H., and Lependu, P. 2010. OntoGrate: Towards automatic integration for relational databases and the semantic web through an ontology-based framework. In: *Int. Journal of Semantic Computing*, 4(01), 123-151.

[13] Guarino, N. 1997. Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. In: *Information Extraction A Multidisciplinary Approach to an Emerging Information Technology*. Springer Berlin Heidelberg, 1997. p. 139-170.

[14] Gruber, T. 1993. A translation approach to portable ontology specifications. In: *Knowledge acquisition*, 5(2), 199-220.

Estudo da aplicação de técnicas inteligentes em mineração de processos de negócio

Alternative Tittle: Study of the intelligent techniques application in business process mining

A.R.C. Maita
Universidade de São Paulo
Av. Arlindo Béttio, 1000
Ermelino Matarazzo
ar.cardenasmaita@usp.br

Prof. Dr. M. Fantinato
Universidade de São Paulo
Av. Arlindo Béttio, 1000
Ermelino Matarazzo
m.fantinato@usp.br

Profa. Dra. S. M. Peres
Universidade de São Paulo
Av. Arlindo Béttio, 1000
Ermelino Matarazzo
sarajane@usp.br

RESUMO

Mineração de processos se situa entre a mineração de dados e aprendizado de máquina, de um lado, e a modelagem e a análise de processos de negócio, de outro lado. Mineração de processos visa descobrir, monitorar e melhorar processos de negócio reais por meio da extração de conhecimento a partir de *logs* de eventos disponíveis em sistemas de informação orientados a processos. Considerando que essas técnicas são, atualmente, as mais aplicadas nas tarefas de mineração de dados de forma geral, seria esperado que elas também estivessem sendo majoritariamente aplicadas nas tarefas de mineração de processos de forma específica, o que não tem sido demonstrado na literatura recente. Este projeto de Mestrado busca compreender porque técnicas inteligentes não têm sido amplamente usadas neste contexto.

Palavras-Chave

Gestão de processos de negócio, mineração de processos, mineração de floxos de processos

ABSTRACT

Mining process lies between data mining and machine learning, on one hand, and business process modeling and analysis, on the other hand. The mining process aims at discovering, monitoring and improving business processes by extracting real knowledge from event logs available in process-oriented information systems. Whereas these techniques are currently the most applied in data mining tasks, it would be expected that they were being mostly applied to process mining tasks as well, which has been not shown in recent literature. This master's project seek to understand why these techniques have not been widely in this context.

SBSI 2015, May 26th-29th, 2015, Goiânia, Goiás, Brazil

Categories and Subject Descriptors

H.4 [**Informaiton Systems Applications**]: Workflow managementDecision support systems; I.2 [**Artificial intelligence**]: Learning

General Terms

Business process mining

Keywords

Business process mining, process mining, workflow mining, workflows mining, mining workflows.

1. INTRODUÇÃO

Gestão de Processos de Negócio (BPM – *Business Process Management*) inclui métodos, técnicas e ferramentas para apoiar o projeto, a realização, o acompanhamento e a melhoria de processos de negócio – ou workflows de negócio – [6]. O ciclo de vida de BPM inclui as fases [7]: *(i)* modelagem de processos; *(ii)* implementação de modelos de processo; *(iii)* execução e administração de instâncias de processo; *(iv)* monitoramento e auditoria de instâncias de processo em execução; e, *(v)* avaliação e melhoria de modelos de processo [7]. Nessa última, o histórico de execução do processo pode ser avaliado visando a sua otimização.

Mineração de dados se refere à extração ou mineração de conhecimento a partir de grandes quantidades de dados[3]. Técnicas de inteligência computacional têm se mostrado bastante eficientes na resolução das tarefas de mineração de dados, pois possuem boa capacidade em lidar com dados imprecisos e incompletos, gerando modelos flexíveis e com graus de precisão elevados. Assim, técnicas de inteligência computacional apresentam-se como alternativas às técnicas exatas, que embora gerem modelos altamente precisos são, muitas vezes, impraticáveis em problemas reais [1].

Na junção das duas áreas — BPM e mineração de dados —, um novo campo de estudo é estabelecido, chamado mineração de processos de negócio[5]. Trata-se de aplicar tarefas de mineração de dados sobre dados provenientes do ciclo de vida de BPM. O objetivo é extrair conhecimento de eventos (por meio de *logs*) provenientes do trabalho realizado nas diferentes fases de um processo de negócio, buscando melhorar tal processo, por meio da descoberta de associações

entre variáveis, de padrões de comportamento ou de desvios de comportamento.

2. APRESENTAÇÃO DO PROBLEMA

Atualmente, cada vez mais eventos de execução são registrados, facilitando a criação de históricos de processos [**?**]. Além disso, existe uma crescente necessidade de melhorar processos em ambientes competitivos e de rápida evolução. Consequentemente, existe interesse na área de mineração de processos [5] por permitir às organizações atender suas necessidades de aprender sobre seus próprios processos [4].

Como a mineração de processos deriva principalmente da mineração de dados, esta área foi melhorada e adaptada para criar técnicas usadas para minerar registros de dados que contém os dados de execução do processo. Esses registros são os *logs* de execução, que são tipicamente armazenados em sistemas de BPM, embora eles também possam ser acessados por meio de outros sistemas relacionados ao processo. Além disso, algoritmos personalizados foram desenvolvidos especificamente para atender às necessidades de especialistas em mineração de processos. Pela análise já realizada em [4], as técnicas de mineração de processos mais empregadas são aquelas que incorporam heurísticas, que são normalmente usadas em mineração de dados para resolução de tipos de problemas mais triviais.

Yue *et al.* [9] descrevem diversos algoritmos e ferramentas de mineração de processos propostos por pesquisadores representativos nessa área. Entre elas, estão: algoritmo para modelar processos usando máquinas de estado; uso de redes neurais artificiais; uso de cadeias de Markov; modelagem de processos com expressões booleanas; tratamento de ruído e estruturas paralelas; os algoritmos α (alpha) e β (beta) para descobrir processos por meio de redes de Petri; métodos MergeSeq, SplitSeq e SplitPar usando grafos de tarefas estocásticas, para tratar tarefas duplicadas; algoritmo para lidar com modelos de workflow de estrutura hierárquica; a ferramenta *Process Miner* que gera modelos de processo por meio da construção de blocos; e métodos para reescrever modelos. Uma desvantagem comum em muitos desses trabalhos é a presença no resultado de valores atípicos e excepcionais (conhecidos como "ruído").

Por meio de uma avaliação exploratória inicial, verifica-se que, embora mineração de processos seja uma junção de BPM e mineração de dados, muitas técnicas e algoritmos ampla e satisfatoriamente usados no contexto geral de mineração de dados são raramente usados no contexto específico de mineração de processos, mais especificamente, técnicas de inteligência computacional. Se pudesse ser verificado que tais técnicas também apresentam bons resultados em mineração de processos, as lições aprendidas com essa verificação representariam uma importante contribuição para os pesquisadores e profissionais da área.

3. PROPOSTA DA SOLUÇÃO

Espera-se que este projeto de Mestrado possa oferecer como resultado, a pesquisadores e profissionais da área de mineração de processos, benefícios para seus trabalhos futuros, como um conjunto de lições aprendidas, sobre a análise dos trabalhos existentes e dos resultados obtidos no estudo de caso. Trata-se de um projeto de pesquisa interdisciplinar que busca contribuir com a área de mineração de processos por meio do oferecimento de uma visão ampla de vantagens e desvantagens da aplicação das técnicas mais modernas de inteligência computacional.

O principal objetivo deste projeto de Mestrado é investigar o cenário de aplicação de técnicas de inteligência computacional no contexto de mineração de processos. Para alcançar esse objetivo é necessário: *(i)* confirmar se técnicas de inteligência computacional realmente são pouco usadas em mineração de processos; *(ii)* analisar os motivos pelos quais essas técnicas não têm sido amplamente usadas em tal contexto, em comparação à área de mineração de dados; e, *(iii)* investigar se essas técnicas têm potencial para apresentar melhores resultados para a mineração de processos comparativamente a outras técnicas tradicionalmente aplicadas.

A metodologia de desenvolvimento compreende: *(i)* realizar uma revisão bibliográfica, um mapeamento sistemático geral, e revisão sistemática focada no uso de redes neurais artificiais (ANN – *Artificial Neural Network*) e máquinas de vetores-suporte (SVM – *Suport Vector Machine*), escolhidas porque elas podem ser aplicadas em diferentes tarefas de mineração de dados[2], e SVM, por exemplo, foi incluída entre os "top 10"algoritmos em mineração de dados de acordo com[8]; *(ii)* investigar as características específicas do contexto de mineração de processos, e de BPM, em relação aos outros contextos genericamente usados em mineração de dados, buscando evidências para a justificativa ou não de uma maior ou menor aplicação de técnicas inteligentes em cada caso; *(iii)* realizar estudos de caso buscando aplicar, em mineração de processos, técnicas inteligentes amplamente aplicadas no contexto geral de mineração de dados, em função das conclusões obtidas das atividades anteriores; e, *(iv)* analisar os resultados obtidos com o estudo de caso visando generalizar as conclusões obtidas para cenários diversos de mineração de processos.

Uma revisão bibliográfica foi realizada envolvendo as principais envolvidas neste projeto: Processos de negócio, mineração de dados, e mineração de processos. Atualmente uma revisão bibliográfica detalhada encontra-se em andamento sobre tópicos específicos a serem usados nos experimentos. Informações adicionais sobre a execução do mapeamento sistemático e a revisão sistemática, mencionados como passos da metodologia, são apresentadas na Seção 4.

Em relação aos estudos de caso, pretende-se comparar os resultados da aplicação de diferentes técnicas selecionadas em um mesmo cenário de processo. Para a avaliação dos resultados, uma técnica de medição de custo/benefício será escolhida, a depender da escolha das técnicas da etapa anterior. Os dados para desenvolver os estudos de caso proveem de *logs* de processos registrados no uso de um sistema de ensino à distância de cursos de especialização oferecidos pela Universidade de São Paulo em conjunto com a Universidade Virtual do Estado de São Paulo.

4. ATIVIDADES REALIZADAS

Durante a condução do projeto foi necessária colaboração dos professores Dr. M. Fantinado (orientador) e Dra. S. M. Peres (co-orientadora), especialistas em Gestão de Processos de Negócio e, Mineração de dados e Inteligência Computacional, respectivamente. Suas opiniões ajudaram a esclarecer as discussões levantadas em cada uma das atividades realizadas.

Dois tipos de revisão da literatura estão sendo tratados neste trabalho. A revisão sistemática, por ser de escopo me-

nor foi focada no uso da técnica de ANN e SVM no contexto de mineração de processos. Os resultados da revisão já foram submetidos na forma de um artigo a um periódico internacional. Por outro lado, o mapeamento sistemático, com um escopo bem maior, ainda se encontra em andamento. Os dois tipos de revisão são necessários como parte dos resultados para alcançar os objetivos de pesquisa deste projeto, não sendo tratadas como simples revisões bibliográficas. O mapeamento sistemático permite confirmar se técnicas que apresentam bons resultados na área de mineração de dados e inteligência computacional estão sendo aplicadas em mineração de processos. O mapeamento sistemático e revisão sistemática foram conduzidas en duas etapas de forma similar: *(i)* identificação e seleção dos estudos primários, foram escolhidas como fontes de dados de pesquisa as bases de Scopus e ISI Web of Science, resultando 3.107 artigos encontrados entre 2003 e 2013; com a aplicação dos critérios de inclusão e exclusão foram selecionados 603 estudos primários; *(ii)* definição de critérios de classificação específicos para cada um dos tipos de revisão.

A revisão sistemática selecionou estudos primários que tratam de ANN e SVM especificamente (11 artigos). Uma avaliação da qualidade dos estudos primários foi realizada. A revisão sistemática, já permitiu verificar que, embora haja interesse científico na área de mineração de processos, pouco tem sido investido especificamente em técnicas como ANN e SVM, apenas 2% de todo o universo de trabalhos identificados na área (603) aplica ANN e SVM. Esse baixo porcentual pode ser decorrência de um possível baixo conhecimento de suas potencialidades para esse tipo de problema. Considerando que a área de mineração de processos envolve uma série de conhecimentos multidisciplinares, isso pode provocar, em muitos casos, a falta de experiência em determinadas áreas de conhecimento. Em relação ao tipo de mineração de processos tratado especificamente quando esses dois tipos de técnicas são usadas, verificou-se que dois estudos se referem ao tipo descoberta, três a extensão, e seis a conformidade. Embora essa classificação tenha sido realizada considerando as definições apresentadas por van der Aalst [5], de fato, a maioria dos estudos primários identificados não se mostraram plenamente aderente à classificação de tipos de mineração de processos usada.

No mapeamento sistemático foram definidos diversos critérios de classificação para as tarefas de mineração de dados, tipo de algoritmo/técnica usado, assim como os tipos de mineração de processo que pretende-se resolver em cada um dos estudos encontrados. A classificação do total de artigos foi finalizada, mas devido à demora na análise, pela abrangência de artigos, não prevista inicialmente, uma atualização (ainda em andamento) para considerar as publicações do ano 2014 foi necessária.

Os resultados parciais do mapeamento já indicam o crescente interesse na área de mineração de processos. Apesar disso, esse interesse ainda segue predominantemente as abordagens tradicionais de mineração de dados, principalmente na escolha de técnicas e algoritmos para tratar os diferentes tipos de mineração de processos. Os resultados confirmam que embora existam trabalhos que tratem as técnicas de inteligência computacional e mineração de dados como, por exemplo, ANN, SVM, algoritmos genéticos, e lógica *fuzzy*; essas não são majoritariamente usadas. Em relação aos tipos de mineração de processos tratados, a maior parte dos estudos primários trata de descoberta de processos. Isso se deve provavelmente ao fato dos outros dois tipos de mineração de processos partirem do uso do modelo de processo, que pode precisar ser primeiramente descoberto.

5. CONCLUSÃO

Este é um projeto interdisciplinar que visa dar uma maior visibilidade de informações a pesquisadores e profissionais interessados em mineração de processos. Este trabalho busca mostrar que o uso de técnicas de inteligência computacional e mineração de dados em geral poderiam oferecer bons resultados também na área de mineração de processos. Espera-se realizar isso por meio de: estudos do estado da arte desta área; estudo e compreensão da razão porque tais técnicas de mineração de dados não estariam sendo amplamente usadas em mineração de processos; casos de estudo com a aplicação dessas técnicas; e, finalmente, proposta de um guia de aplicação de técnicas inteligentes para mineração de processos.

A revisão sistemática realizada para identificar e avaliar os trabalhos que propõem o uso de redes neurais artificiais ou máquinas de vetores-suporte no contexto da mineração processo demonstra que, embora haja interesse científico na área de mineração de processos, pouco tem sido investido especificamente nesse tipo de técnica. Além disso, com o mapeamento sistemático – ainda em andamento – já é possível verificar que outras técnicas e abordagens inteligentes amplamente usadas em mineração de processos, em geral, são pouco usadas no contexto de mineração de processos.

Os próximos passos deste trabalho permitirão comprovar possíveis hipóteses levantadas ao finalizar a análise do mapeamento sistemático; assim como generalizar conclusões obtidas em cenários diversos de mineração de processos.

6. REFERÊNCIAS

[1] J. Abonyi, B. Feil, and A. Abraham. Computational intelligence in data mining. *Informatica*, 29(1):3–12, June 2005.

[2] S. da Costa Crtes, R. M. Porcaro, and S. Lifschitz. Mineração de dados–funcionalidades, técnicas e abordagens. *PUC-RioInf, Rio de Janeiro. In the Brazilian goverment using Credit Scoring Leonardo sales*, May 2002.

[3] H. Jiawei and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publisher, United States of America, SF, 2006.

[4] A. Tiwari, C. J. Turner, and B. Majeed. A review of business process mining: State-of-the-art and future trends. *Business Process Management Journal*, 14.

[5] W. van der Aalst. *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag Berlin Heidelberg, Netherlands, AZ, 2011.

[6] W. van der Aalst, ter Arthur Hofstede, and M. Weske. Business process management: A survey, June 2003.

[7] M. Weske. *Business Process Management*. Springer-Verlag Berlin Heidelberg, Germany, Potsdam, 2006.

[8] W. Xindong et al. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, October 2007.

[9] D. Yue et al. A review of process mining algorithms. *International Conference on Business Management and Electronic Information*, 5:181–185, May 2011.

Quais são as questões em BPM na perspectiva Brasileira?

Alternative Title: What are the issues in BPM in Brazilian perspective?

Valdemar T. F. Confort
Univ. Federal do Estado do Rio de Janeiro
Av. Pasteur, 458, Urca
Rio de Janeiro, RJ, Brasil
valdemar.confort@uniriotec.br

Flávia Maria Santoro
Univ. Federal do Estado do Rio de Janeiro
Av. Pasteur, 458, Urca
Rio de Janeiro, RJ, Brasil
flavia.santoro@uniriotec.br

ABSTRACT

Business Process Management (BPM) is a subarea within Information Systems where international publications identifies its evolution of. Other publications identify practical challenges in several perspectives. Our proposal is to identify, in Brazilian perspective, the evolution of the academic interest and the practical challenges of the national organizations. The expected result is to answer our research question: What are the issues in BPM in Brazilian perspective? In addition, we expect to contribute with an instrumentation that can be applied in future evaluations, following the same process of this research.

Categories and Subject Descriptors

General Terms

Management, Measurement.

Keywords

Business Process Management.

1. INTRODUÇÃO

Dumas et al [1] definem Gestão de Processos de Negócio (GPN, ou BPM do inglês *Business Process Management*) como a arte e ciência de supervisionar como o trabalho é realizado na organização para assegurar resultados consistentes e obter vantagem das oportunidades de melhoria. A GPN combina conhecimentos da tecnologia da informação e das ciências da administração e as aplica em processos de negócios operacionais [1]. Processos de negócio, por sua vez, são coleções inter-relacionadas de eventos, atividades e pontos de decisões que envolvem uma quantidade de atores e objetos e que levam, coletivamente, a um resultado que é um valor para pelo menos um cliente [2] [3]. A pesquisa neste campo resultou numa ampla quantidade de métodos, técnicas e ferramentas para apoiar o desenvolvimento, implantação, gerenciamento e análise de processos de negócios operacionais [2]. A GPN interessa a diversos grupos numa organização, desde administradores responsáveis pela empresa até os participantes do processo que executam atividades no dia a dia.

SBSI 2015, May 26–29, 2015, Goiânia, Goiás, Brazil.

Tanto a academia quanto as organizações possuem interesse mútuo em GPN, sendo que os pesquisadores reconhecem os desafios práticos e concordam que há um aumento de complexidade e de escopo dos processos nas organizações [2] [4] [5] [6] [7]. Recker apresenta também algumas evidências relevantes das preocupações das organizações [8]. Primeiro, GPN é um desafio para os executivos mais experientes [9]; segundo, em 2009, WinterGreen previu que o mercado de soluções de GPN triplicaria entre 2009-2014, com um incremento de US$ 6,2 bilhões de dólares [10]; finalmente, Indulska et al relatam que as organizações ainda lidam com as fases iniciais e triviais de descobrir e documentar os seus processos de negócio [11].

Algumas iniciativas contribuem para condensar a evolução do conhecimento na área de GPN. A partir de uma perspectiva internacional e acadêmica, Aalst discute a evolução dentro da Conferência Internacional de BPM [12]. No nível prático e restrito ao universo da Austrália, uma série de trabalhos correlacionados foi apresentada, por exemplo, [13] [14] [15]. No Brasil, podemos citar trabalhos no nível prático [16], e iniciativas na academia que contribuem para o estado da arte [17]. No entanto, não existe uma visão consolidada do cenário de GPN no Brasil, seja no estado da arte ou da prática. Sem esta visão não é possível entender os desafios que se colocam diante das organizações e nem como estas podem se posicionar perante o mercado internacional.

Assim, alguns fatos ensejam o presente trabalho. O primeiro fato é que há reconhecida **relevância** da Gestão de Processos de Negócio tanto pela indústria quanto pela academia. O segundo é que as conferências nacionais fornecem subsídios para uma análise consolidada que identifique o estado da arte. O terceiro é o de que há linhas de base para que comparações possam ser feitas tanto em relação ao estado da arte, vide trabalho de Aalst [12], quanto em relação ao estado da prática [13]. E, finalmente, o fato da existência desta lacuna de que não existe um trabalho consolidador que permita iniciar tanto uma discussão do ponto de vista evolutivo quanto uma discussão comparativa, seja do ponto de vista acadêmico quanto do ponto de vista prático.

Este cenário e fatos motivam a contribuir com a realização deste diagnóstico, isto é, consolidar o estado da arte e o estado da prática no Brasil. Reconhecer esse cenário é o primeiro passo para evoluí-lo. Desta forma a questão a ser tratada neste trabalho é: **Quais são as questões em BPM na perspectiva Brasileira?**

2. TRABALHOS RELACIONADOS

Em 2003, Aalst, Hofstede e Weske publicam uma pesquisa que condensa os conhecimentos em GPN [2]. No artigo os autores contextualizam historicamente o surgimento dos sistemas de gestão

de processos de negócio (BPMS, do inglês Business Process Management System), apresentam os conceitos fundamentais e o ciclo de vida da Gestão de Processos de Negócios, discutem acerca de método e modelagem e expõem as tecnologias e padrões emergentes à época.

Em 2013, Aalst publica uma pesquisa mais abrangente [12]. Neste artigo, todos os itens da pesquisa de 2003 sofrem um maior adensamento. Além disso, Aalst apresenta duas perspectivas de classificação: 20 casos de usos para demonstrar "como, onde e quando" GPN pode ser aplicada; e 6 preocupações chaves da área. Aalst ainda apresenta os artigos da Conferência Internacional em BPM segundo estas classificações. Desta forma, ele apresenta a evolução dos artigos tanto em relação aos casos de uso quanto em relação aos conceitos chaves. Para isso, ele etiqueta cada artigo segundo uma das classificações e determina a frequência relativa com que os assuntos são pesquisados. A Tabela 1 exemplifica o resultado desta análise das preocupações chaves por ano [12].

Tabela 1. Análise da frequência relativa das preocupações chaves por ano conforme Conferência Internacional [12].

Ano	Linguagem de Modelagem de Processo	Infraestrutura de Execução de Processo	Análise de Modelo de Processo	Mineração de Processo	Flexibilidade de Processo	Reuso de Processo
2000	0.355	0.161	0.290	0.000	0.161	0.032
2003	0.325	0.200	0.250	0.050	0.075	0.100
2004	0.286	0.238	0.238	0.143	0.048	0.048
2005	0.288	0.231	0.212	0.058	0.096	0.115
2006	0.154	0.308	0.288	0.096	0.077	0.077
2007	0.387	0.097	0.194	0.194	0.065	0.065
2008	0.324	0.108	0.297	0.135	0.081	0.054
2009	0.148	0.111	0.370	0.222	0.037	0.111
2010	0.240	0.240	0.200	0.160	0.000	0.160
2011	0.143	0.171	0.200	0.314	0.000	0.171
Média	0.265	0.187	0.254	0.137	0.064	0.093

Já em relação ao estado da prática, diversas também são as publicações. Induska et al sugerem uma abordagem multi-metodológica para pesquisar quais são as maiores questões em GPN na perspectiva australiana [13] (Figura 1).

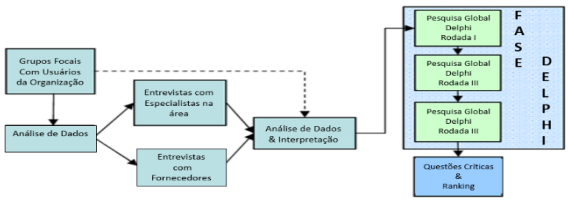

Figura 1. Abordagem multi-metodológica [13].

No primeiro artigo desta linha de pesquisa, é apresentado o resultado do estudo feito com a metodologia de grupos focais [13]. No segundo artigo relata-se o resultado de entrevistas realizadas com especilistas [14]. Por fim, no último artigo, os autores relatam o resultado de uma pesquisa feita com os fornecedores de soluções em GPN [15].

No Brasil, a área de Gestão de Processos de Negócio vem sendo estudada no campo dos Sistemas de Informação cuja conferência principla é o Simpósio Brasileiro de Sistemas da Informação que, em 2014, realizou sua décima edição [18]. Dentro deste congresso foi abrigada a oitava edição do *Workshop* de Gestão de Processos de Negócio [24], que em 2015, participa do evento como uma trilha. Não identificamos publicações que apontem, de forma consolidada quais são as questões principais da academia.

Em relação à prática, existem publicações que permitem observar o cenário brasileiro. Em 2008, Paim et al publicam um artigo sobre um estudo de caso acerca da estruturação de um escritório de Gestão de Processos [16]. Além do estudo de caso o artigo contribui, em termos de generalização, ao apresentar um *benchmark* acerca dessa implantação em cinco organizações. Santos et al, em 2011 [19], apresenta um estudo empírico, de abordagem qualitativa, sobre a adoção de GPN em quatro organizações públicas. O artigo apresentou o resultado acerca de três questões de pesquisa: Quais são os objetivos da iniciativa de BPM? Quais são as abordagens metodológicas que estão sendo usadas? Que resultados e benefícios têm sido obtidos pelas iniciativas de BPM? Em 2014, um novo estudo, com as mesmas questões é publicado mas, com três organizações privadas [20].

3. ENFOQUE DE SOLUÇÃO

Considerando que um projeto de pesquisa é realizado para determinar como a pesquisa será conduzida, o presente trabalho utilizou como referência as fases apresentadas por Recker [21] (Figura 2). A fase exploratória visa a construção de um entendimento sobre o fenômeno de interesse. A fase de racionalização tem como objetivo dar sentido às coisas que envolvem o problema de interesse. A fase de validação visa submeter a teoria ou suposição a avaliações rigorosas.

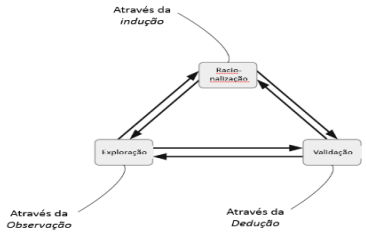

Figura 2. Projeto de Pesquisa segundo Recker [21]

A presente pesquisa propõe, em fase exploratória, realizar uma pesquisa de literatura de conferências nacionais sobre GPN. Nessa linha temos como referências imediatas o trabalho de Aalst [12]. Na fase de racionalização, será realizada uma pesquisa com abordagem semelhante aos trabalhos de Indulska [13] e Paim [16]. Justifica-se a semelhança de abordagem, pois é desejável que os resultados sejam passíveis de comparação. Particularmente em relação ao trabalho de Paim, possibilitará entender se houve ou não evolução do cenário apresentado em 2007.

Na fase de validação, pretende-se submeter a abordagem da fase de racionalização a uma avaliação criteriosa. No trabalho de Indulska [13] os autores realizaram uma pesquisa com Grupos Focais. Nesse sentido, pretende-se desenvolver o estudo de forma iterativa de forma que os riscos sejam minimizados e para que seja possível avaliar se, de fato, identifica-se as questões pertinentes à realidade nacional.

Resumindo, o enfoque de solução tem duas perspectivas: uma acadêmica e uma prática. A abordagem será construída de tal forma que possa ser identificada como um instrumento que possibilite a qualquer tempo a realização de um novo diagnóstico.

4. CONCLUSÃO

Esta pesquisa parte destes dois fatos expostos: i) relevância reconhecida da área de GPN e; ii) identificação a importância de trabalhos que permitem uma análise consolidada tanto do estado da arte quanto do estado da prática. Assim, a proposta é fornecer uma visão da perspectiva nacional sobre o tema nesses dois aspectos: arte e prática. Este artigo apresentou alguns trabalhos relacionados que servirão como guias desta pesquisa.

Esta pesquisa está em fase inicial. Na perspectiva acadêmica, foi realizada a coleta dos artigos do Workshop de Gestão de Processos de Negócio. Após a coleta, foi aplicada uma análise em relação às preocupações chaves de GPN. Conforme planejado, foi possível comparar o resultado consolidado desta primeira análise em nível nacional com o trabalho realizado apresentado por Aalst [12]. Para exemplificar, incluímos a Figura 3 que apresenta um dos resultados observados.

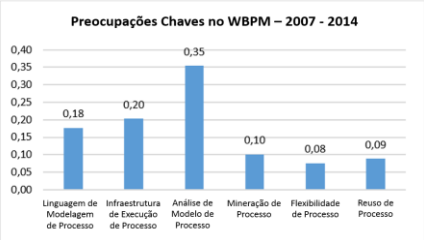

Figura 3. Importância relativa das preocupações chaves no Workshop de BPM (WBPM) Brasil.

Os próximos passos serão a continuidade desta análise do cenário acadêmico e a evolução para a perspectiva da prática, que ainda está em fase de planejamento. A fase prática exige um planejamento maior tendo em vista as diversas dificuldades esperadas inerentes a este tipo de pesquisa.

A principal contribuição esperada é a resposta à questão de pesquisa. A academia possuirá uma análise para balizar novas pesquisas bem como permitirá a comparação do nível nacional com o internacional. Por sua vez, as organizações e praticantes terão à disposição uma pesquisa que retrate os desafios do ponto de vista nacional. Desta forma, academia e indústria poderão trocar experiências. Os desafios da indústria poderão ser enfrentados com conhecimentos já consolidados na academia ou, vice-versa, poderão contribuir com novas questões de pesquisa. Finalmente, outra contribuição é em relação a possibilidade de que a abordagem realizada possa ser instrumentalizada de tal forma que, no futuro, haja a realização de um novo diagnóstico. Neste sentido, a abordagem de *Design Research* [21] pode se mostrar apropriada para tal objetivo.

5. REFERÊNCIAS

[1] M. Dumas, M. La Rosa, J. Mendling and H. A. Reijers, Fundamentals of Business Process Management, Berlin: Springer, 2013.

[2] W. Aalst, A. Hotsfede and M. Weske, "Business Process Management: a survey," in *BPM 2003*, Berlin, 2003.

[3] W. Aalst, "Business Process Management Demystified: A tutorial on models, systems and standrds for workflow management," in *Lecture Notes in Computer Science*, Berlin, Germany, 2004.

[4] P. Akhavan, M. Jafari and A. Ali-Ahmadi, "Exploring the interdependency between reengineering and information technology by developing a conceitual model," *Business Process Management Journal,* v.12, no.4, pp.517-534, 2006.

[5] R. Paim, H. Caulliraux and R. Cardoso, "Process Management Tasks: a conceptual and practical view," *Business Process Management Journal,* 2007.

[6] P. Fettke, P. Loos and J. Zwicker, "Business Process Reference Models: Survey and Classification," in *BPM 2005 Workshops*, 2005.

[7] P. Harmon, Business Process Change: A Manager´s Guide to Improving, Redesigning, and Automating Processes, San Francisco: Morgan Kaufmann, 2003.

[8] J. Recker, Evaluations of Process Modeling Grammars, v.71, Berlin: Springer, 2011.

[9] Gartner Group, "Gartner Group: Leading in Times of Transition: The 2010 CIO Agenda. Exp Premier Report," Gartner, Connecticut, 2010.

[10] WinterGreen Research, Inc., "WinterGreen Research: Business Process Management (BPM) Market Opportunities, Strategies, and Forecasts, 2009 to 2015.," WinterGreen Research, Lexington, 2009.

[11] M. Indulska, S. Chong, W. Bandara, S. Sadiq and M. Rosemann , "Proceeding of the 17th Australasian Conference on Information Systems.," in *Australasian Association for Information Systems*, Adelaide, 2006.

[12] W. Aalst, "Business Process Management: A Comprehensive Survey," *ISRN Software Engineering,* v.13, 2013.

[13] M. Indulska, S. Chong, W. Bandara, S. Sadiq and M. Rosemann, "Major issues in business process management: an Australian perspective," in *17th Australasian Conference on Information System, Australasian Association for Information System*, Adelaide, 2006.

[14] W. Bandara, M. Indulska, S. Chong and S. Sadiq, "Major issues in business process management: An expert perspective," *BPtrends,* pp. 1-8, 2007.

[15] S. Sadiq , M. Indulska, W. Bandara and S. Chong, "Major issues in business process management: a vendor perspective," in *Proceedings of the 11th PACIS*, Auckland, New Zealand, 2007.

[16] R. Paim, V. Nunes, B. Pinho, F. Baião, C. Cappelli and F. M. Santoro, "Structuring a Process Management Office: Case Studies in Brazilian Companies," *Business Process Management Journal,* 2011.

[17] "I Workshop Brasileiro em Gestão de Processos em Negócios," Gramado, RS, 2007. "II Workshop Brasileiro em Gestão de Processos em Negócios," Vila Velha, ES, 2008. "III Workshop Brasileiro em Gestão de Processos em Negócios," Fortaleza,CE, 2009. "IV Workshop Brasileiro em Gestão de Processos em Negócios," Maraba, PA, 2010. "V Workshop Brasileiro em Gestão de Processos em Negócios," Salvador, BA, 2011. "VI Workshop Brasileiro em Gestão de Processos em Negócios," São Paulo, SP. "VII Workshop Brasileiro em Gestão de Processos em Negócios," 2013. "VIII Workshop Brasileiro em Gestão de Processos em Negócios," Londrina, 2014.

[18] "Simpósio Brasileiro de Sistemas da Informação," Londrina, 2014.

[19] H. M. Santos, A. F. Santana, G. Valença and C. F. Alves, "Um Estudo Exploratório sobre adoção de BPM em Organizações Públicas," *V Workshop Brasileiro em Gestão de Processos em Negócios,* 2011.

[20] C. Rômulo, F. Carvalho, A. Queiroz, R. Monteiro and R. José, "Uma Análise Exploratória sobre Adoção de BPM em Organizações Privadas," *VIII Workshop de Gestão de Processos de Negócio,* 2014.

[21] J. Recker, Scientific Research in Information Systems: A Beginner's Guide, Springer Science & Business Media, 2012.

Regras de Negócio para Cidadãos: Compreensão e Comunicação

Business Rules for Citizens: Comprehension and Communication

Matheus Masseron Sell
Programa de Pós-Graduação em Informática
Núcleo de Pesquisa e Inovação em CiberDemocracia
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro RJ, Brasil
matheus.sell@uniriotec.br

Renata Araujo
Programa de Pós-Graduação em Informática
Núcleo de Pesquisa e Inovação em CiberDemocracia
Universidade Federal do Estado do Rio de Janeiro (UNIRIO),
Rio de Janeiro RJ, Brasil
renata.araujo@uniriotec.br

ABSTRACT
This research aim at proposing the use of declarative languages (SBVR) as a way to communicate public services process rules to citizens, allowing rules understanding, and consequently, the discussion and collaboration among citizens towards service improvements by communicating their experiences and opinions about the service.

RESUMO
Essa pesquisa tem o objetivo de utilizar linguagens declarativas (SBVR) como uma forma de comunicar regras de processos de serviços públicos para cidadãos, permitindo a compreensão de regras, e consequentemente, a discussão e colaboração entre cidadãos para a melhoria de serviços pela comunicação das suas experiências e das suas opiniões sobre o serviço.

Categorias e Descritores de Assunto
D.2.9 [**Knowledge Representation Formalisms and Methods**]: Representations (procedural and rule-based)

Termos Gerais
Linguagens, Fatores Humanos

Palavras Chave
Regras de negócio, cidadão, SBVR, serviços públicos

1. INTRODUÇÃO
A participação dos cidadãos nos processos de prestação de serviços públicos pode ser promovida pelo uso de modelos de processos de negócio como instrumento para comunicar ao cidadão como um serviço é oferecido. A simples apresentação de um modelo de processos para um cidadão, no entanto, não é suficiente para seu entendimento, pois artefatos técnicos podem não ser facilmente compreendidos por cidadãos (Araujo et al., 2014).

Regras representam as limitações impostas aos processos de prestação de um serviço, tornando-se um dos elementos fundamentais e de interesse de conhecimento pelos cidadãos. A apresentação das regras aos cidadãos e seu entendimento pode vir a favorecer a participação cidadã em serviços públicos, uma vez que ao compreender as regras, os cidadãos passam a compreender os serviços prestados pelas instituições e, consequentemente, podem se comunicar e discutir sobre eles. Nesse sentido, surge a questão desta pesquisa: Como prover aos cidadãos a compreensão das regras de negócio associadas ao processo de prestação de um serviço público, permitindo o diálogo e discussão sobre as mesmas entre os cidadãos e entre os cidadãos e as instituições públicas provedoras do serviço?

A proposta deste trabalho é especificar, implementar e avaliar uma solução computacional de apoio à participipação de cidadãos à compreensão e discussão sobre regras de negócio. Para isso devem ser estudados três aspectos: i) o uso de uma linguagem de descrição de regras compreensível ao cidadão; ii) a comunicação entre cidadãos a partir das regras; e iii) formas de possibilitar a proposição de mudanças nas regras destes processos pelos cidadãos.

2. ePARTICIPAÇÃO EM SERVIÇOS PÚBLICOS
A Democracia Eletrônica pode ser entrendida dentro de um contexto de participação civil em democracias mediado pelas tecnologias de informação e comunicação (TICs). A CiberDemocracia envolve a criação de novos processos, estimula o uso de tecnologias de interação social e promove a participação e transparência de ações (Pimentel et al., 2011).

Araújo e Taher (2014) propõem cinco níveis de participação cidadã em assuntos públicos por meio de TICs. Cada um desses níveis compreende um conjunto de requisitos de colaboração, transparência e memória social voltados à construção de ferramentas de apoio à eParticipação: Provisão de informações e serviços; Coletar opinião pública e o uso dessa informação para tomada de decisão política; Transparência e responsabilidade, este nível é responsável por uma maior permeabilidade governamental; Democracia deliberativa onde são estabelecidos processos e mecanismos para discussões para tomada de decisão que passam por um sistema deliberativo; Democracia direta e co-design onde as discussões não passam por um sistema deliberativo.

3. MODELAGEM DE PROCESSOS PARA A PARTICIPAÇÃO CIDADÃ
Uma informação desejável em organizações, incluindo as públicas, é a informação sobre os processos que são executados ou geridos por elas. Toda a empresa, independentemente do

tamanho deve lidar com um conjunto de processos de negócio (Dumas et al., 2012). BPM (Business Process Management) é a área que estuda como o trabalho ocorre em uma organização para permitir resultados consistentes e ter vantagens de oportunidades de melhorias.

Alguns trabalhos propuseram o uso de modelos de processos como instrumento de transparência organizacional e base para o diálogo com cidadãos. Dirr (2011) realizou um estudo sobre conversas sobre processos de prestação de serviços públicos com o auxílio de uma ferramenta computacional. O objetivo da proposta foi promover a interação entre cidadãos e prestadores de serviço. Gomes (2011) realizou um estudo sobre a compreensão de regras em processos de prestação de serviços públicos utilizando recursos de animação. Engiel (2012) propôs uma abordagem para o projeto de características de entendimento para modelos de processo de prestação de serviços públicos visando os usuários externos à organização.

Os resultados destas pesquisas apontam que as regras de negócio nos processos tendem a ser elementos de interesse e de discussão pelos cidadãos. Também como conclusão destas pesquisas, destaca-se a sugestão de estudar formas de facilitar a compreensão sobre as regras dos processos, sua discussão e até proposição de mudanças pelos cidadãos.

3.1. COMPREENSIBILIDADE DE MODELOS DE PROCESSOS

A modelagem de processos auxilia a compreender processos junto com as pessoas que estão envolvidas com ele no dia a dia. Uma linguagem de modelagem de processos foca no aspecto das mudanças contínuas dos objetos do processo, isso quer dizer que em um modelo pode ser explicitado como os atores executam diferentes atividades com o decorrer da execução, como artefatos mudam entre diferentes etapas do processo, e como regras podem oferecer diferentes restrições em momentos distintos dentro de uma mesma execução do processo.

Fahland et al. (2009) afirma que a questão de compreensibilidade é um pilar na busca de linguagens de modelagem de processos. Essa busca pela questão de compreensibilidade estende-se a questões do paradigma utilizado para modelar, como, por exemplo, o uso de linguagens imperativas ou declarativas. Modelos imperativos propõem uma modelagem estruturada. Explicitam uma ordem ou uma sequência entre as atividades, enquanto modelos declarativos permitem especificar regras entre as atividades (Schaid et al., 2014).

Em um modelo imperativo deve ser encontrada uma trajetória contínua que representa o comportamento do processo (Fahland, 2009). Estabelecer uma relação sequencial de informação é mais fácil sobre a base do modelo que é criado em uma linguagem de modelagem de processos que é relativamente mais imperativa por natureza.

Fahland (2009) afirma que linguagens de modelagem de processos declarativas focam na lógica que governa a relação geral entre objetos e ações de um processo. Essa relação pode ser arbitrária, não necessita ser contínua, ela deve somente descrever a lógica do processo. Dessa forma, qualquer comportamento que satisfaça o modelo é viável no processo. Estabelecer uma relação circunstancial de informação é mais fácil sobre a base do modelo que é relativamente mais declarativo por natureza. Pode-se supor que processos que utilizam informações circunstanciais são melhor resolvidos utilizando linguagens declarativas. Em um processo público, o cidadão possui interessse nas circunstancias nas quais ele está encaixado, ou seja, quais regras podem permitir ou impedir que ele execute alguma ação.

3.1.1. SBVR

SBVR (Semantics of Business Vocabulary and Rules) é uma linguagem para a modelagem de um texto formal e estruturado. A linguagem faz parte da OMG Model Driven Architecture. O documento de especificação da linguagem, OMG (2008), define SBVR como uma linguagem que suporta vocabulário de negócio, regras de negócio e lógica de primeira ordem. Por se tratar de uma notação que se baseia em regras ela pode ser considerada uma linguagem de natureza declarativa. Em SBVR o domínio de negócio é representado pelo vocabulário de SBVR (*SBVR vocabulary*). Inglês Estruturado (*Structured English*) é proposto como como uma notação para as regras de SBVR.

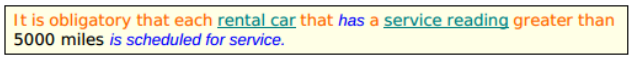

Figura 1 - Exemplo de notação de uma regra em SBVR, Object Management Group (2008)

4. COMPREENSÃO E COMUNICAÇÃO DE REGRAS PELOS CIDADÃOS

Regras de negócio delimitam o que um cidadão pode um não pode fazer em um processo, assim como pode determinar pré-requisitos das atividades do processo. Esta pesquisa propõe que, a respeito de regras em processos de prestação de serviços públicos, um cidadão possa:

- Compreender as regras- Neste trabalho foi escolhido SBVR para expressar regras de negócio para o cidadão devido à sua característica de representação textual e a necessidade do uso de uma modelo formal. O equilíbrio entre uma representação formal e a compreensibilidade para o cidadão é necessário pois um modelo formal pode auxiliar na automatização da execução do processo.

- Comunicar-se - Cidadãos devem ser capazes de comunicar experiências e opiniões sobre um processo e suas regras considerando essa necessidade de co-participação na compreensão de regras de negócio. A motivação da comunicação pode ser vista pelos seguintes exemplos: Cidadãos podem demonstrar insatisfação em relação a uma etapa de algum serviço e oferecer sugestões de melhorias. Cidadãos também podem ter dúvidas relativas às regras, essas dúvidas podem ser solucionadas por cidadão que tiveram problemas semelhantes.

- Co-editar regras - Nos níveis quatro e cinco de participação democrática (Araujo e Taher, 2014) deve haver alguma forma para que o cidadão participe da tomada de decisão. A edição participativa pode ser uma forma dos cidadãos participarem com deliberações de como processos devem ser executados e criados.

Um artefato computacional será desenvolvido para apoiar os aspectos da solução. O projeto do artefato será uma aplicação móvel e inicialmente o foco será nas questões de interação, apresentação de regras, discussão e proposição de mudanças. Em trabalhos futuros será considerada a execução e a adaptação de processos junto ao cidadão.

5. ABORDAGEM METODOLÓGICA

O objetivo desta pesquisa é estudar se a representação de regras por meio de adaptações na SBVR pode ser viável para auxiliar a compreensão e facilitar a troca de informações sobre serviços públicos por cidadãos.

Um estudo de caso deve ser aplicado a fim de demonstrar o quais são os pontos positivos e negativos da aplicação de SBVR em um ambiente colaborativo para a colaboração, a comunicação e a co-edição de processos. O estudo será focado na compreensão de processos modelados em SBVR que terá uma característica confirmatória buscando entender se o uso dessa linguagem com o apoio de uma ferramenta computacional facilita a compreensão do cidadão. Esse estudo busca também analisar como os cidadãos se comunicam utilizando-se de uma ferramenta que estutura a conversa, seguindo paradigmas propostos na área de Sistemas Colaborativos. Foram realizados dois estudos exploratórios. Um estudo avaliou a compreensão de SBVR e outro observou a interação entre pessoas durante a discussão sobre regras.

A hipótese avaliada nesse trabalho é: SE forem apresentadas as regras de negócios em SBVR ao cidadão ENTÃO ele poderá compreender, comunicar e sugerir alternativas às regras. Um grupo de participantes será escolhido e serão propostas algumas tarefas e discussões para que os participantes da pesquisa completem ou participem. Será analisada a quantidade de tarefas concluídas na ferramenta para avaliar a hipótese.

Será verificada a quantidade de tarefas concluídas para analisar uma hipótese verdadeira. O falseamento será atingido caso: a maioria dos participantes não consiga completar as tarefas, ou a maioria dos participantes não consiga compreender as regras de negócio expressas, ou a maioria dos participantes complete as tarefas de forma incorreta. O estudo da discussão será falseado caso a apresentação dos serviços em regras de negócio não mantenha as manifestações direcionadas no serviço ou caso as conversas entre os envolvidos não forneça insumos para identificação de melhorias do serviço.

6. CONCLUSÃO

Entre as contribuições deste trabalho espera-se na área de BPM ser a identificação de características que podem ser encontradas no uso de regras de negócio expressas em SBVR para o entendimento de um processo por parte de cidadãos. Em Sistemas Colaborativos deve ser estudada uma forma para promover a troca de informações sobre regras de processos de negócio. Em Democracia Eletrônica espera-se possibilitar uma forma de discussão e compreensão sobre regras de negócio, um elemento de serviços públicos e favorecer o debate entre cidadãos veiculado por um artefato computacional.

7. REFERÊNCIAS

[1] ANTTIROIKO, A. V., “Building Strong E-Democracy - The Role of Technology in Developing Democracy for the Information Age”, Communications of the ACM, 2003

[2] ARAUJO, R. M. ; CAPPELLI, C. ; ENGIEL, P. . Raising Citizen - Government Communication with Business Process Models. In: Cemal Dolicanin ; Ejub Kajan; Dragan Randelovic; Boban Stojanovic;. (Org.). Handbook of Research on Democratic Strategies and Citizen-Centered E-Government Services. 1ed.-: IGI Global, , v. 1, p. 1-389, 2014

[3] ARAUJO, R. M.; TAHER, Y.: Refining IT requirements for Government-Citizen co-participation support in public service design and delivery, Conference for Democracy and Open Government CeDEM14, 2014

[4] CAPPELLI, C., LEITE, J. C. S. P, “Transparência de Processos Organizacionais”, Dumas, M., La Rosa , M., Mendling, J., Reijers , H. A., “Fundamentals of Business Process Management ”, 2012

[5] DIRR B., ARAUJO R., CAPPELLI C., "Conversas sobre processos de prestação de serviços públicos", Dissertação de Mestrado, UNIRIO, Rio de Janeiro, 2011

[6] DUMAS, M., LA ROSA , M., MENDLING, J., REIJERS , H. A.: Fundamentals of Business Process Management Berlin: Springer, 2012.

[7] ENGIEL, P.; ARAUJO, R.; CAPPELLI, C.; Projetando o entendimento de modelos de processos de prestação de serviços públicos; 2012. 113 f. Dissertação (Mestrado em Informática) – Centro de Ciências Exatas e Tecnologia, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro. 2012.

[8] FAHLAND, D., LBKE, D., MENDLING, J., REIJERS, H., WEBER, B., WEIDLICH, M., ZUGAL, S.: Declarative versus imperative process modeling languages: The issue of understandability; Springer Berlin Heidelberg. 2009

[9] OBJECT MANAGEMENT GROUP; Semantics of Business Vocabulary and Business Rules (SBVR); disponível em http://www.omg.org/spec/SBVR/1.2/PDF/; acesso em 19 de março de 2015, 2008

[10] PIMENTEL, M.; FUKS, H.; Sistemas Colaborativos, Elsevier, Rio de Janeiro, 2011

[11] REIJERS, H. A., SLAATS, T., STAHL, C., Declarative Modeling - An Academic Dream or the Future for BPM?, Lecture Notes in Computer Science, Volume 8094, 2013

[12] RODRIGUES, R. A., AZEVEDO, L. G., REVOREDO, K., LEOPOLD, H., “A Tool to Generate Natural Language Text from Business Process Models ”, CBSOFT, 2014

[13] SCHAIDT, S., SANTOS, E. A. P., LOURES, E. D. F. R., E BUSETTI, M. A., “Gerenciamento de Processos de Negocio Não Estruturados: Implementacão baseada em LTL,” VIII Simpósio Brasileiro de Sistemas de Informação (SBSI 2012) VI Workshop de Gestão de Processos de Negócio (WBPM 2012), 2012.

[14] SILVA, S. P.; Graus de participação democrática no uso da Internet pelos governos das capitais brasileiras, Opinião Pública, v. xi. n.2 450-468, 2005

Sincronização Automática dos Artefatos de Processos de Negócio: Métodos e Aplicações

Automatic Synchronization of BPM Description Artifacts: Methods and Applications

Raphael de A. Rodrigues
Programa de Pós-Graduação em Informática (PPGI/UNIRIO)
Av. Pasteur, 456, Urca, Rio de Janeiro, RJ, Brasil
raphael.rodrigues@uniriotec

Leonardo G. Azevedo
Programa de Pós-Graduação em Informática (PPGI/UNIRIO)
Av. Pasteur, 456, Urca, Rio de Janeiro, RJ, Brasil
IBM Research
Av. Pasteur, 136, Botafogo, Rio de Janeiro, RJ, Brasil
azevedo@uniriotec.br, lga@br.ibm.com

Kate Revoredo
Programa de Pós-Graduação em Informática (PPGI/UNIRIO)
Av. Pasteur, 456, Urca, Rio de Janeiro, RJ, Brasil
katerevoredo@uniriotec.br

RESUMO

A validação de modelos de processos de negócio é executada por analistas de sistemas. Estes podem ler facilmente o modelo, porém não detêm conhecimento aprofundado do negócio. Em contrapartida, especialistas de domínio conhecem o 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 boa comunicação entre especialistas e analistas. Este trabalho propõe uma metodologia para geração de textos a partir de modelos e vice-versa. A metodologia será avaliada em cenários reais, através de empresas e estudos de caso.

Palavras-Chave

Geração de Linguagem Natural, modelos de processo, texto em linguagem natural, BPMN

ABSTRACT

System 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. On 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 good communication between business specialits and system 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 methodology will be evaluated in real scenarios, through companies and case study.

SBSI 2015, May 26th-29th, 2015, Goiânia, Goiás, Brazil

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; I.2.7 [**Computing Methodologies**]: ARTIFICIAL INTELLIGENCE Natural Language Processing [Language generation, Machine translation]

General Terms

Design, Documentation, Languages

Keywords

Natural language generation, process models, natural language text, BPMN

1. INTRODUÇÃO

Diversas organizações utilizam notações para descrever seus processos de negócio, como a BPMN [2]. A notação BPMN foi desenvolvida e padronizada pelo *Object Management Group*. 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 [5], [1]. 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.

Este trabalho propõe a criação de uma abordagem bilateral que seja capaz de manter ambos os artefatos (modelos e textos) atualizados automaticamente.

Este artigo está dividido nas seguintes seções. Na seção 2 é apresentada a revisão bibliográfica do tema. Na seção 3 é apresentada a proposta deste trabalho para a manutenção automática dos artefatos de processos de negócio. Na seção 4 são apresentados os passos empregados na execução deste trabalho. Finalmente, na Seção 5 são apresentadas as conclusões do trabalho e propostas de trabalho futuro.

2. REVISÃO BIBLIOGRÁFICA

Diversas técnicas de verbalização foram propostas para solucionar o problema de inconsistência, permitindo que as organizações evitem trabalho redundante. Estas técnicas permitem um mapeamento direto de modelos conceituais para linguagem natural [7]. Em pesquisas anteriores, elaboramos as primeiras propostas para a transformação de modelos de processos para textos em linguagem natural [4], [9]. 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*).

Este trabalho propõe a criação de uma abordagem bilateral capaz de interpretar modelos de processos de negócio, escritos em inglês ou em português, mapeando-os para linguagem natural. As alterações detectadas no texto produzido (ou no modelo) deverão ser automaticamente refletidas a fim de manter ambos os artefatos sincronizados.

Em particular, o trabalho deverá responder a seguinte pergunta de pesquisa: *Como as organizações podem manter ambos artefatos, modelos e textos descritivos de processo, automaticamente sincronizados, tendo em vista que o processo manual está sujeito a falhas na consistência da informação?*

3. TÉCNICA PARA MANUTENÇÃO AUTOMÁTICA DE ARTEFATOS

Nesta seção é apresentada a técnica para manutentção automatica de artefatos.

3.1 Visão Geral da Técnica

A figura 1 detalha a visão geral da arquitetura proposta para a solução bidirecional. A ideia principal da arquitetura é manter um conjunto de correlações entre elementos do modelo de processo e do texto gerado. Baseando-se na construção da técnica de geração anterior [9] (passos 1 e 2), tal correlação pode ser produzida automaticamente. O passo 1 é responsável por gerar estruturas intermediárias que armazenam informações sobre o modelo. O passo 2 é responsável por ler tais estruturas e mapeá-las para sentenças em linguagem natural. Se um usuário aplicar alterações no texto em linguagem natural (passo 3), estas alterações serão analisadas pelo `Analisador de Alterações`. As ligações entre os elementos do modelo e sentenças em linguagem natural permite detectar quais elementos do modelo foram afetados pela alteração textual e como estas alterações impactam no modelo original. No passo 4, as alterações no texto são refletidas no modelo de processo. O acionamento do `Analisador de Alterações` está vinculado à detecção de alterações em os ambos artefatos, textos e modelos.

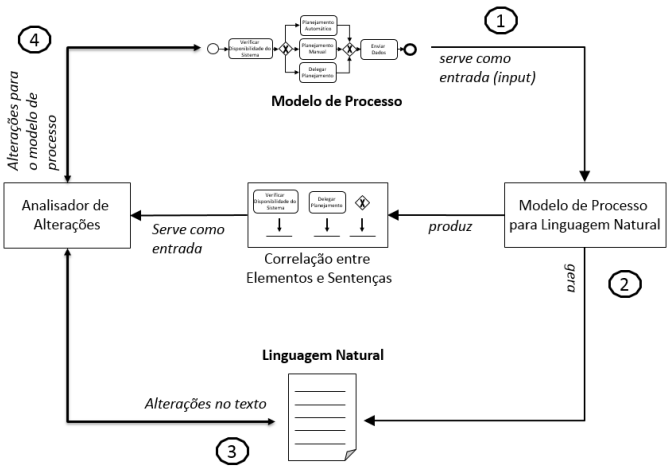

Figura 1: Visão geral da solução.

Além do projeto dos componentes principais, os objetivos foram estendidos 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. 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 [6], [3].

2. **Independência da Linguagem para Modelagem**: Na indústria diversas linguagens de modelagem são utilizadas. Frequentemente, encontram-se exemplos de modelos elaborados em BPMN, EPCs e UML. Reconhecendo esta característica, será definida 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.*, [6], [3]). Será definido 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, a abordagem bidirecional poderá ser empregada para diferentes linguagens de modelagem.

3. **Solução *User-friendly***: É importante definir uma técnica que possa ser eficientemente utilizada na prática. Desta forma, a abordagem desenvolvida será integrada em um editor que esteja disponível ao público, como por exemplo, o editor online Signavio[1]. Acredita-se que este plug-in será de grande serventia tanto para a indústria, quanto para a academia.

4. METODOLOGIA E CONTRIBUIÇÕES

A metodologia a ser aplicada será baseada na proposta de métodos e aplicações destes, além da especificação de experimentos e pesquisas qualitativas [10] para avaliação da abordagem proposta. Detalhando esta metodologia temos:

1. Levantamento e estudo aprofundado sobre o estado da arte nas seguintes áreas: Geração de Linguagem Natural, Modelagem de Processos, Revisão de Teorias e Refinamento de Modelos.

2. Monitoramento de novos trabalhos e propostas publicados em congressos, periódicos e simpósios que estejam no escopo deste trabalho.

3. Desenvolvimento de métodos abordando as propostas do presente trabalho.

4. Avaliação dos métodos desenvolvidos, inicialmente com dados artificiais, o que permitirá um experimento controlado. Posteriormente, serão utilizados dados reais, através de um estudo de caso [8]. O estudo de caso contará com empresas e universidades da Áustria e do Brasil. Tal estudo será possível graças a colaborações estabelecidas entre os países citados. O estudo permitirá avaliar o comportamento da ferramenta sob condições reais e que não foram definidas a priori. Serão aplicados questionários desenvolvidos pelo aluno (questionário este, que ainda está em elaboração) nas universidades e empresas parceiras, com a ajuda de nossos colaboradores.

5. Avaliação e publicação dos resultados parciais e finais.

O processo de experimentação já foi iniciado pelos autores, sendo classificado como *"true experimental design"* [8], pois técnicas de randomização foram aplicadas durante o processo de seleção das amostras para garantir que os grupos do experimento fossem similares entre si. No total, 73 indivíduos (estudantes e profissionais de TI) foram selecionados para participar do experimento. Este primeiro experimento teve como objetivo responder a seguinte pergunta de pesquisa: *"Qual representação pode ser entendida mais facilmente pelo usuário, modelos de processo ou textos descritivos, de acordo com a experiência em modelagem de processo?"*. Os resultados estão sendo publicados e estarão disponíveis, caso forem aceitos, no SBES 2015[2].

O trabalho produz 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.

[1] www.signavio.com

[2] Maiores detalhes podem ser obtidos em http://cbsoft.org/sbes2015

5. CONCLUSÃO

Este trabalho 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: (i) 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; (ii) Apresentar a correlação entre elementos do modelo de processo e elementos da linguagem natural; (iii) Desenvolvimento de uma abordagem bidirecional, eficiente e efetiva, integrada entre modelos e textos.

Por fim, este trabalho 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 e vice-versa. Pretende-se, através dos algoritmos desenvolvidos, a criação de uma ferramenta capaz de gerar um texto em ambas as línguas, português e inglês.

6. REFERÊNCIAS

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

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

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

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

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

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

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

[8] J. Recker. *Scientific research in information systems: a beginner's guide*. Springer, 2012.

[9] R. Rodrigues, L. Azevedo, K. Revoredo, and H. Leopold. A tool to generate natural language text from business process models. In *CBSoft 2014 - Tool Session*, sep 2014.

[10] J. Wainer. Métodos de pesquisa quantitativa e qualitativa para a ciência da computação. *Atualização em informática*, pages 221–262, 2007.

Uma abordagem baseada em análise ontológica para alinhamento entre conceitos de negócio e modelos conceituais

Alternative Title: An approach based on foundational ontology to align business concepts and conceptual models

Patricia Merlim L. Scheidegger
Universidade Federal do Rio de Janeiro
Departamento de Ciência da Computação iNCE
Programa de Pós-graduação em Informática
patricia.merlim@ufrj.br

Maria Luiza Machado Campos
Universidade Federal do Rio de Janeiro
Departamento de Ciência da Computação iNCE
Programa de Pós-graduação em Informática
mluiza@ppgi.ufrj.br

RESUMO
Esta pesquisa propõe uma abordagem de análise ontológica aplicada à construção de definições com o objetivo de criar mecanismos que permitam gerar definições consistentes de conceitos e representá-los de forma que sejam inteligíveis e utilizáveis no âmbito computacional.

Palavras Chave
Definições, Modelagem Conceitual, Ontologia de Fundamentação

ABSTRACT
In this research we describe an approach through ontological analysis applied at the construction of business definitions in order to create mechanisms that can lead to consistent definitions and represent them in a way that can be understandable and useful to IT artifacts.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *abstracting methods, dictionaries, linguistic processing*.

General Terms
Design, Theory.

Keywords
Definitions, Conceptual Modeling, Foundational Ontology

1. INTRODUÇÃO
O alinhamento entre a Tecnologia da Informação (TI) e o negócio vem ganhando importância nas organizações e é considerado fator essencial para alcance de seus objetivos estratégicos. Equipes de TI se esforçam para alcançar a missão de alinhar tecnologia ao negócio e procuram resgatar e registrar os conceitos de negócio.

Entretanto, apesar destes esforços, prevalece uma distância entre o conceito em sua natureza e o conceito tal qual implementado nos sistemas de informação.

Frequentemente, artefatos de controle terminológico (glossários, tesauros), mostram-se pouco sistemáticos na construção das definições associadas aos termos.

Durante o ciclo de desenvolvimento de sistemas não é comum um esforço inicial direcionado ao levantamento e entendimento dos principais conceitos sob um ponto de vista abrangente, contextual. A visão costuma ser focada na forma como os recursos computacionais irão captar, disponibilizar e manter os dados associados ao conceito.

Este trabalho tem como objetivo a proposta de uma abordagem para construção de definições mais consistentes visando preservação de semântica nas transposições dos conceitos entre artefatos de TI durante o ciclo de desenvolvimento de sistemas.

2. APRESENTAÇÃO DO PROBLEMA
É de fundamental importância que conceitos sejam bem definidos e que sua definição seja compreensível até por quem não é familiarizado com seus diferentes contextos. É essencial compreender um conceito, de que forma ele pode ser melhor definido, independente de implementação. É importante refletir sobre a informação que queremos representar e buscar resgatar e imprimir sua semântica nesta representação [9].

O mundo real é subjetivo e as definições e interpretações também podem ser. Segundo Thiry-Cherques [12] "todo esforço deve ser feito para que a definição facilite o entendimento completo do que são os atributos incluídos no conceito".

Em um glossário podemos encontrar definições incipientes, e principalmente, inconsistentes entre si dentro do mesmo contexto. Em geral os glossários não são construídos por especialistas em conceitualização e mesmo quando o são, nem sempre se mostram adequados, para serem utilizados pelos profissionais de TI nas etapas posteriores do ciclo de desenvolvimento de sistemas.

As etapas do ciclo de desenvolvimento de sistemas, em especial as iniciais de levantamento de requisitos e análise geram perda de conhecimento se forem conduzidas de forma incorreta. Há um esforço grande em equalizar entendimento entre especialistas do domínio e profissionais de TI e é nestas etapas a maior concentração de informação desencontrada, chamada por Wanderley e Silveira [14] de *gap* semântico.

O processo de modelagem conceitual por sua vez envolve percepções humanas naturalmente subjetivas e compreende duas atividades principais: levantamento dos conceitos do universo de discurso que será modelado e representação dos conceitos levantados de acordo com a gramática de uma linguagem de modelagem [3]. O processo de modelagem conceitual é similar ao de tradução, pois é preciso entender os conceitos expressos em linguagem natural e então representá-los em outra linguagem, a de modelagem [3]. Um modelo conceitual rico em semântica deve prover os elementos necessários para representar a realidade da forma mais fidedigna possível, e de modo que seja compreensível [3], independente de implementação.

A linguagem de modelagem deve ter os construtos essenciais para representar a especificação do modelo, com suas entidades e relações. Se a linguagem for pouco expressiva, boa parte da conceitualização não será representada na linguagem [7]. Quanto mais conhecemos um determinado domínio e quanto mais precisos formos em sua representação, maiores são as chances de termos sistemas computacionais consistentes e compatíveis com a realidade deste domínio. Ter uma representação precisa de uma conceitualização se torna ainda mais importante e crítico quando queremos integrar diferentes modelos que compartilham um mesmo domínio. Vemos atualmente nas empresas um descasamento entre conceitos definidos em glossários e a representação destes conceitos em modelos conceituais.

2.1 Trabalhos Relacionados

Medeiros [10] apresenta análise comparativa entre tesauros e ontologia de fundamentação. A análise de Medeiros observa que as definições de termos do tesauro, em geral, "não estão em consonância com a estrutura hierárquica estabelecida, possibilitando a ocorrência de falhas na representação de um domínio". Outro trabalho relacionado [11] propõe uma metodologia de análise comparando a estrutura interna de definições com a estrutura interna de modelos relacionais pré-estabelecidos por cada tipo. Estes modelos foram criados usando como base a ontologia de fundamentação BFO (Basic Formal Ontology). A BFO foi concebida estritamente para o campo das ciências naturais, em especial a área de biomédicas e limita o uso da metodologia em outros domínios. Castro, Baião e Guizzardi [3] apresentaram uma proposta de modelagem conceitual baseada em análise linguística seguindo uma teoria e classificação de tipos semânticos, tais como *Animate, Human, Parts, Inanimate* e usando uma linguagem de modelagem cujos construtos são baseados em ontologia de fundamentação.

Nossa proposta vai além da validação de definições. Ela trata da sistematização para construção de definições baseada em uma ontologia de fundamentação independente de domínio e voltada para modelos conceituais. A modelagem conceitual tende a se beneficiar ao utilizar definições de conceitos de negócio mais completas e consistentes.

3. PROPOSTA DE SOLUÇÃO

Para alinhamento entre conceitos de negócio e modelos conceituais propomos uma abordagem baseada no processo sistemático de construção de definições a partir da análise ontológica do conceito.

Através do processo de análise ontológica acreditamos ser possível olhar o conceito a partir de sua natureza, resgatando sua essência. O resultado beneficia uma representação com semântica agregada, possibilitando assim que o conceito seja explicitado de forma mais precisa, para que possa ser melhor compreendido e também mais adequado, para ser utilizado como base construída a partir dele, pelos demais artefatos de TI.

Em uma abordagem de análise ontológica identificamos as metapropriedades [5] associadas ao conceito, tais como identidade, rigidez e dependência, e buscamos representá-lo explicitando estas metapropriedades através dos construtos de uma ontologia de fundamentação. Ontologias de fundamentação "são sistemas de categorias filosoficamente bem fundamentados e independentes de domínio, que têm sido utilizados com sucesso para melhorar a qualidade de linguagens de modelagem e modelos conceituais" [6].

Guizzardi [7] propõe uma ontologia de fundamentação denominada UFO (Unified Foundational Ontology) que foi desenvolvida com o intuito de prover uma fundamentação ontológica para linguagens de modelagem conceitual. A UFO tem sido aplicada com sucesso para avaliar linguagens de modelagem conceitual assim como prover semântica do mundo real para seus elementos utilizados em uma modelagem [6].

Nossa pesquisa utilizará os princípios e construtos da UFO para, a partir da natureza do conceito, resgatar os elementos essenciais que devem estar presentes em uma definição de forma a torná-la consistente e reutilizável pelos artefatos de TI, a exemplo de modelos conceituais. Acreditamos que o alinhamento entre as definições de negócio e os artefatos de TI será promovido naturalmente dado que tanto as definições quanto os modelos conceituais derivados a partir delas se utilizariam da mesma base teórica.

Considerando as categorias básicas da UFO, dependendo da natureza, cada definição tem aspectos distintos. Se algo é definido como *tipo* entende-se que seus subtipos fazem parte de sua extensão. As *fases* por sua vez são instanciadas a partir de uma propriedade intrínseca. Quando identificamos um *papel* precisamos também identificar qual a relação que o torna um papel [1]. Se para cada categoria existem elementos essenciais, que na nossa proposta serão aplicados nas definições dos conceitos, estes mesmos elementos, quando aplicáveis a modelagem conceitual, naturalmente estarão presentes.

Como um exemplo de análise ontológica temos o *papel*. Os papéis geralmente são exercidos por seres ou coisas que possuem identidade (*tipos*) e tem a metapropriedade dependência. Esta dependência é existencial em uma relação material, ou seja, para ser um papel precisa necessariamente da relação com outro indivíduo para existir. O papel de estudante, exercido por uma pessoa, depende da matrícula em uma instituição de ensino. A partir dos princípios apresentados pela UFO, entendemos que o papel é antirigido, o que quer dizer que os seres ou coisas que exercem os papéis não necessariamente o exercem durante toda a sua existência.

Em resumo, para que a definição de um papel esteja clara e completa é preciso identificar os seguintes elementos: o *tipo* que exerce o papel e a relação que o origina. Com estes elementos presentes na definição, o analista de TI obtem a indicação da importância de explicitar a relação que origina o papel que está sendo modelado.

Este trabalho ainda está em desenvolvimento e se propõe a estabelecer diretrizes para construção de definições de conceitos de naturezas diferentes como *papéis, relações parte-todo, tipos, fases* e *eventos*.

4. PROJETO DE AVALIAÇÃO DA SOLUÇÃO

A abordagem de solução será aplicada em artefatos reais corporativos. Utilizaremos glossários desenvolvidos por áreas de negócio e os modelos conceituais mantidos por TI que representam os principais conceitos existentes nestes glossários. Como validaremos sobre definições já existentes desenvolveremos no escopo deste trabalho um processo de validação das definições a partir das diretrizes de nossa proposta. Nossa abordagem será empírico-analítica para inicialmente evidenciarmos o problema. Posteriormente, ao propor melhorias nas definições e relacioná-las a mudanças nos modelos poderemos verificar seu impacto consultando os analistas.

As etapas de validação estão em planejamento. O escopo deve abranger validação de definições; validação de modelos conceituais; relatórios com gaps semânticos; relatórios com sugestão de definições; questionários para feedback dos analistas quanto aos resultados apresentados. Pretendemos trabalhar sobre domínio chave da empresa para o qual os resultados teriam uma abrangência mais significativa e além dele conceitos de pelo menos duas áreas de negócio diferentes.

5. ATIVIDADES JÁ REALIZADAS

Iniciamos com uma revisão de literatura sobre o tema definições nos seus diferentes contextos, tanto pela Ciência da Informação [10,11,12,13] quanto pela Ciência da Computação. Na Ciência da Computação estudamos o papel das definições no ciclo de desenvolvimento de sistemas[14]; ontologia de fundamentação como base para conceitualização [2,4,5,6] e os aspectos essenciais do processo de modelagem conceitual [3,9]. Pelo referencial teórico ser ontologia de fundamentação, em especial a UFO, foi realizado um estudo detalhado de suas metapropriedades e categorias [1,4,7,8].

Em seguida iniciamos um estudo empírico utilizando um glossário corporativo e um recorte de modelo conceitual que abrangia alguns dos conceitos presentes no glossário. Analisamos a definição de alguns conceitos chave e a modelagem conceitual destes conceitos, aplicando a sistematização das metapropriedades visando verificar a viabilidade da proposta.

A fase atual é de finalização da diretriz de construção de papéis e de sua correlação com os elementos da modelagem conceitual. Pretendemos estender estes procedimentos às demais categorias de definições mencionadas anteriormente. A proposta foi defendida em junho de 2014.

6. CONCLUSÃO

Os conceitos nascem na área de negócio de uma organização e são representados nos artefatos de TI. O alinhamento entre negócio e TI é fator crítico de sucesso nas organizações

A partir da análise ontológica baseada em ontologia de fundamentação podemos garantir que aspectos essenciais à natureza do conceito estejam presentes na sua definição. Um glossário com definições bem consistentes pode levar para a modelagem características que em geral não são garantidas de estarem evidentes em regras de negócio. Através da análise ontológica podemos identificar lacunas que poderiam ser preenchidas e obter modelos mais expressivos, assim como glossários que sirvam ao negócio, mas também garantam a coerência e captura do conceito. Desta forma podemos agregar semântica ao conceito antes que o mesmo siga para os demais artefatos que são construídos durante o ciclo de desenvolvimento de sistemas evitando assim problemas tais como ambiguidades e inconsistências que possam ocorrer devido à falta de clareza do conceito.

7. REFERÊNCIAS

[1] Almeida, J., Guizzardi, G. 2008. *A Semantic Foundation for Role-Related Concepts in Enterprise Modelling*, IEEE International EDOC Conference. Proceedings of the 12th IEEE International EDOC Conference. IEEE Computer Society Press

[2] Baião, F., Souza, J., Azevedo L., Nunes, V., Cappelli, C., Santoro, F., Siqueira, S., Castro, L., e Lopes, M. 2009. *Pesquisa e Prática em Ontologias e em Modelagem de Processos de Negócio aplicados à Arquitetura de Informações*. II Seminário de Pesquisa em Ontologias no Brasil

[3] Castro, L., Baião, F., Guizzardi, G. 2010. *A Linguistic Approach to Conceptual Modeling with Semantic Types and OntoUML*. EDOCW 2010. pp. 215-224

[4] Guarino, N. 1998. *Formal Ontology and Information Systems.* Proceedings of FOIS'98. Trento, Italy, June 6-8. Amsterdan, IOS Press, pp. 3-15

[5] Guarino, N., Welty,C. 2004. *An Overview of OntoClean.* S.Staab, R.Studer (eds.). Handbook on Ontologies. Springer Verlag, pp. 11-159

[6] Guizzardi, G., Falbo, R.A. e Guizzardi, R.S.S. 2008. *A Importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software.* IEEE Latin America Transactions, vol.6, no. 3.

[7] Guizzardi, G. 2005. *Ontological Foundations for Structural Conceptual Models.* CTIT.

[8] Guizzardi, G., Wagner, G. 2008. *What's in a Relationship: An Ontological Analysis*, 27th International Conference on Conceptual Modeling (ER'2008), Barcelona, Spain, Lecture Notes in Computer Science., vol. 5231, pp.83 – 97

[9] Kent, W. 2010. *Data and Reality*, 1st Books Library, 2nd edition

[10] Medeiros, J. 2011. *Tesauros Conceituais e Ontologias de fundamentação: Análise comparativa entre as bases teórico-metodológicas utilizadas em seus modelos de representação de domínios*. Tese (Mestrado em Ciência da Informação) – Universidade Federal Fluminense, Niterói

[11] Seppälä, S. 2012. *Constraintes sur la sélection des informations dans lês definitions termninogrraphiques: vers des modèles relationnels génériques pertinent*. Thèse de doctorat: Univ. Genève, no. FTI 12

[12] Thiry-Cherques, A. R, 2012. *Conceitos e Definições*. ISBN 9788522511365. FGV.

[13] Walton, D. and Macagno, F. 2009. *Classification and Ambiguity: The Role of Definition in a Conceptual System*. Studies in Logic, Grammar and Rhetoric16.29. pp. 245-264

[14] Wanderley, F. and Silveira D.S. 2012. *A Framework to Diminish the Gap between the Business Specialist and the Software Designer*. Eighth International Conference on the Quality of Information and Communications Technology. Lisbon. pp. 199-204

Uma solução de BI em Tempo Real para o Ambiente de Computação em Nuvem

Alternative Title: A Real-Time BI solution on Cloud Computing Environment

André Eduardo Bento Garcia
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Avenida Pasteur 458 – Urca – Brasil – Rio de Janeiro - RJ
andre.garcia@uniriotec.br

Astério Kiyoshi Tanaka
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Avenida Pasteur 458 – Urca – Brasil – Rio de Janeiro - RJ
tanaka@uniriotec.br

Fernanda Araújo Baião
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Avenida Pasteur 458 – Urca – Brasil – Rio de Janeiro - RJ
fernanda.baiao@uniriotec.br

RESUMO

Computação em nuvem se tornou uma realidade nos dias atuais. As empresas que não têm a TI como seu negócio fim podem usufruir dos serviços na nuvem. Em paralelo a esta realidade, um ambiente de inteligência de negócios (BI) em tempo real é uma necessidade latente para o acirrado mercado dos negócios, onde a análise e tomada de decisões sobre um repositório integrado de informações permanentemente atualizado pode oportunizar ganhos e redução de custos. Existem desafios de desempenho para a construção de um ambiente de BI em base de dados relacional em tempo real sobre infraestruturas de nuvem. O presente trabalho introduz as primeiras ideias na direção de uma proposta de uma solução para BI em tempo real implementado em infraestrutura de computação em nuvem com foco no ganho de desempenho de processamento.

Palavras-Chave

BI em tempo real, computação em nuvem, data warehouse, banco de dados relacional.

ABSTRACT

Cloud computing has become a reality in current IT environments. Companies that do not have IT as their business purpose can make use of cloud services. In parallel to this fact, a real-time business intelligence (BI) environment is a latent need for organizations, where the analysis and decision-making on top of an integrated repository of information that is constantly updated may generate opportunity gains and reduce costs. There are performance challenges for the construction of a real-time BI environment on top of a cloud infrastructure. This paper introduces the first ideas towards an efficient service for real-time BI in the cloud.

Categories and Subject Descriptors

H.2.4 [**Database Management**]: Systems – distributed databases, query processing, relational database.

General Terms

Measurement, Performance, Experimentation.

Keywords

Real-Time BI, cloud computing, data warehouse, relational data base.

SBSI 2015, May 26–29, 2015, Goiânia, Goiás, Brazil.

1. INTRODUÇÃO

Computação em nuvem se tornou uma realidade nos últimos anos, com a promessa de entregar recursos de software e hardware sob demanda, suportada por uma infraestrutura de rede [4]. Este é um ambiente propício para implementação de Inteligência de Negócio em Tempo Real (RTBI) que se destaca como um consumidor de recursos significativos para responder às necessidades empresariais. Tal panorama motivou o presente estudo que visa relacionar estes dois temas e apresentar uma solução tecnológica para atender a tal expectativa. Esta seção introdutória trata da conceitualização e definições da fundamentação teórica do tema.

Business Intelligence (BI) define conceitos e métodos para melhorar tomadas de decisões empresariais, utilizando sistema de apoio baseados em fatos [12]. Em geral, BI é composta por uma base de dados que tem como características dados históricos que estão consolidados, limpos, tratados e armazenados um repositório centralizado chamado de *data warehouse* [7]. As fontes dos dados para carga no *data warehouse* são geralmente bases de dados heterogêneas, em geral de sistemas com características operacionais OLTP [10] e [7]. Uma vez que estes dados estão disponíveis no *data warehouse*, podem ser consultados por ferramentas como relatórios analíticos e gerenciais, OLAP e mineração de dados. Uma etapa importante da construção de *data warehouse* é o ETL (Extração, Transformação e Carga), que consiste de rotinas/funcionalidades utilizadas para efetuar a carga de dados a partir de uma base de dados operacional [10].

RTBI – BI em tempo real é uma definição que contempla BI no contexto de acesso a dados históricos, assim como dados operacionais recentemente atualizados, quase em linha com as transações. O objetivo é alcançar a menor latência possível, ou seja, a menor diferença entre o tempo em que o dado chegou à base de dados operacional e está disponível para análise no *data warehouse* [9]. A RTBI visa entregar não somente a informação de cunho estratégico com base em dados de longo prazo como a BI tradicional, mas também entregar dados mais atuais possíveis para uma rápida tomada de decisão operacional, por vezes, em tempo quase real [6] e [16]. Dentre as soluções presentes na literatura para alcançar a implementação de uma RTBI, a proposta de particionar o conjunto de dados, como, por exemplo, dados históricos e dados recentes, apresentada em [7], demonstra resultados satisfatórios [12], [6] e [9].

Computação em nuvem compreende serviços de hardware e software que podem ser acessados via Internet. Representa um modelo para permitir o acesso conveniente, sob demanda, pela rede para um *pool* compartilhado de recursos computacionais configuráveis (por exemplo, servidores, armazenamento, aplicações e serviços) que podem ser rapidamente provisionados e liberados com um esforço mínimo de gerenciamento ou interação

com o provedor de serviço [4] e [8]. Neste cenário, é possível ao usuário requerer um ou mais serviços quando necessitar, tendo como forma de tarifação a cobrança somente dos serviços que foram utilizados, também conhecido como *pay-and-go*. Um pool de recursos está pronto para uso e sua localização física não é de conhecimento do usuário final, bastando o usuário saber que existe um serviço e que este está disponível. Esta característica beneficia fornecedores do serviço em nuvem que gerenciam seus recursos de forma dinâmica. Outra característica relevante no serviço em nuvem é a disponibilização de recurso de forma elástica ou escalável, ou seja, o recurso pode ser estendido ou retraído sempre que necessário [2]. Estes serviços ou componentes são agrupados nas abstrações a seguir [4] e [8]: o SaaS (Software como serviço) disponibiliza acesso a software para usuários e organizações sob demanda, remotamente, propiciando redução no custo de manutenção, atualização, instalação e licenças para uso; o PaaS (Plataforma como serviço) provê serviço computacional que possibilita a organização ou usuário realizar o desenvolvimento de aplicativos na nuvem; Por fim, o IaaS (Infraestrutura como serviço) provê disponibilidade de ativos físicos de rede, poder computacional e de armazenamento.

Estas abstrações são implementadas em 4 tipos de nuvem a saber:

I - Nuvem pública, o modelo mais comum, o qual disponibiliza recursos de software e hardware para o público em geral. II - nuvem privada, impede que os serviços sejam compartilhados com outros usuários senão determinada organização. III - nuvem comunitária, esta abordagem estabelece que somente um subconjunto de usuários utilizem determinados serviços. IV - nuvem híbrida, mistura entre os tipos de nuvens citados.

BI na Computação em nuvem encontra um potencial de recursos e oportunidades, que permite a exploração de características inerentes à Computação em nuvem como eficiência, flexibilidade e escalabilidade, redução de custos entre outros. Percebe-se que esta mudança de paradigma está se tornando uma realidade para as empresas [1], [11], [15] e [3].

2. APRESENTAÇÃO DO PROBLEMA

RTBI proporciona um diferencial na tomada de decisão em relação ao tempo de resposta a problemas de negócios, possibilitando o alcance de oportunidades e vantagens competitivas às empresas. BI em nuvem é uma tendência para onde as empresas estão aportando seus recursos. Logo, seria natural pensar que RTBI também utilize Computação em nuvem. Contudo, essa mudança de paradigma envolve alguns desafios, entre os quais a questão do desempenho é um fator crítico. É provável que a solução em RTBI na nuvem não alcance, com as mesmas configurações de hardware, o desempenho obtido pelas soluções *in-house*, devido às características inerentes a cada solução [14].

Na solução *in-house,* o ambiente passa por controles rígidos sobre fontes de origem de degradação de performance, como, a latência da rede, por exemplo. Este fato deve ser observado quando da adoção do ambiente em nuvem, logo, o item performance é o fator motivador do presente trabalho. A questão a ser respondida é se o desempenho das consultas em uma solução RTBI em base de dados relacional na nuvem é tolerável. Tendo com falseamento a identificação de que ocorreu queda acentuada no desempenho das consultas.

A limitação ao modelo relacional se deve às técnicas de particionamento utilizadas na proposta, que são aplicáveis a tabelas relacionais. Ademais, o modelo relacional é predominante nas soluções de BI corporativos, foco desta pesquisa.

3. TRABALHOS RELACIONADOS

Inmon e outros [7] propõem uma arquitetura de ciclo de vida do dado no DW, a qual sugere que os dados do ambiente operacional sejam imediatamente atualizados no DW; estes dados por sua vez são organizados segundo seu tempo de existência.

Dehne e outros [5] propõem o CR-OLAP, sistema OLAP em tempo real baseado em nuvem. Esta solução é implementada em estruturas de dados em árvore, que permite alcançar baixa latência, possibilitando consultas e inserções de dados concorrentemente em paralelo. Difere da proposta deste artigo por não ser uma solução para banco de dados relacional.

Ainda, Pereira e outros [12] propuseram uma solução para DW em tempo real, implementada com particionamento e distribuição de dados em uma infraestrutura *in-house*, efetivando a arquitetura proposta do ciclo de vida do dado no DW [7]. A proposta do presente artigo é baseada nesta solução para nuvem.

4. PROPOSTA DE SOLUÇÃO

Este projeto propõe a implementação de uma solução RTBI no ambiente em nuvem para base de dados relacional. Esta solução implementará estratégias de particionamento e distribuição de dados, para carga e consulta em tempo real, investigando a viabilidade desta solução em relação ao desempenho das consultas. Um dos principais focos de pesquisas nas soluções RTBI em ambiente tradicional é a identificação de soluções que possibilitem a consulta aos dados em concorrência com as atualizações. Uma das alternativas para implementação do RTBI tradicional é a divisão/particionamento do conjunto de dados de acordo com sua "idade". Por, exemplo, dados mais recentes são carregados em determinada(s) partição(ões) e dados mais antigos em outra(s). Isso proporciona a divisão de um problema grande, em problemas menores, permitindo abordagens adequadas a cada situação [6], [7], [12] e [9].

No presente trabalho, a técnica de particionamento que será utilizada é a de fragmentação horizontal derivada. Trata-se do agrupamento de linhas de dados de uma relação, obedecendo critérios definidos com base em outra relação. Os critérios mais comumente utilizados são particionamento por *Range*, *Round-Robin*, entre outros. O acesso à base de dados é monitorado por uma aplicação que identificará qual porção da base de dados deverá ser acessada, redirecionando o acesso para tal agrupamento observando o critério configurado.

A base de dados será organizada por fragmentos históricos e fragmento em tempo real. O fragmento em tempo real receberá os dados atualizados na base de dados operacional, e também não terá nenhum tratamento de performance como índices ou visões materializadas, uma vez que o objetivo desta partição são as inserções oriundas do OLTP. Quanto menos tratamento ou transformações nesses dados, menor será o tempo de carga na base de dados do DW.

É preciso ressaltar que tanto os fragmentos históricos e em tempo real terão a mesma estrutura física, diferenciando-se pelo volume de dados armazenados, onde a porção dos fragmentos históricos será significativamente maior do que a porção em tempo real. Esta identidade visa manter a transparência na aplicação, onde não haverá a preocupação de reescritas de consultas por apresentar diferentes estruturas.

Kimball define um modelo de dados largamente utilizado na indústria para organizar os dados fisicamente em um DW conhecido como modelo esquema Estrela ou Dimensional [11]. Este modelo é determinado por tabelas Dimensões e Fatos; a primeira armazena informações descritivas utilizadas para realização de filtros e agrupamentos de Fatos, já a segunda

armazena as métricas ou medidas que estão relacionadas a um evento, ou seja, o que deseja ser computado e analisado.

Na tabela Fatos reside o maior volume de tuplas e este se relaciona às diversas Dimensões, sendo grande responsável pelo trabalho de melhoria de desempenho, e, por este motivo, em geral, é a tabela candidata ao particionamento. Já as Dimensões têm como característica possuir um número bem menor de tuplas, e são replicadas nos fragmentos de dados histórico e de tempo real.

Uma vez determinado o critério de particionamento, a distribuição/replicação dos dados se darão na condição de que as bases de dados sejam independentes, ou seja, cada nó será responsável pelo armazenamento, processamento e gerenciamento do seu conjunto de dados recebido ou consultado em seu domínio. Essa independência é chamada de arquitetura *shared nothing*, que tem como característica a inexistência de compartilhamento de processador e memória, seja principal ou secundária.

A solução tem como elo central um *middleware*, um nó mestre, que é o responsável por recepcionar as consultas e atualizações do *host* cliente e submetê-las aos respectivos nós escravos, assim como consolidar os diversos resultados enviados por estes nós, submetendo a resposta final ao *host* cliente.

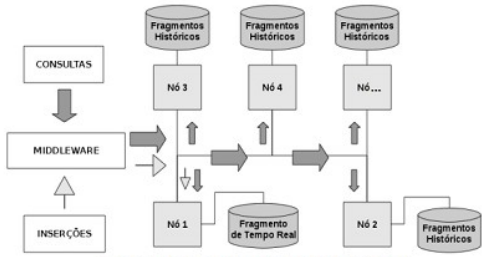

Figura 1 : Arquitetura proposta para o DW no ambiente de BI em tempo-real na nuvem. Adaptado de [12]

A Figura 1 ilustra a arquitetura proposta: O Nó 1, onde reside o fragmento em tempo real, receberá demandas de consultas e inserções, já os demais nós, onde residem os fragmentos históricos, receberão somente demandas de consultas.

É de se esperar que ocorra a degradação da performance do nó em tempo real, o qual receberá as inserções e não está otimizado para não afetar o tempo de carga. De fato, é necessário o estabelecimento de um limite aceitável de tempo de resposta sob este nó; uma vez ultrapassado este limite, será necessária a transferência dos dados deste fragmento para os fragmentos históricos.

5. PROJETO DE AVALIAÇÃO DA SOLUÇÃO

Será realizado método de avaliação quantitativo no qual, por intermédio de experimentos, serão aferidos os tempos das consultas. Será utilizado o *Star Schema Benchmark*, que permite simular as principais características de um cenário *data warehouse* em tempo real, que corresponde a cargas e consultas de dados simultaneamente [13]. Os experimentos serão executados em ambiente em nuvem a ser escolhido entre os diversos fornecidos pelo mercado. Serão executados experimentos em diversos cenários a título de exemplo podemos citar por variação de volume das cargas de dados em tempo real, variação no volume de massa de dados legados e variação do número de nós. As consultas propostas no *Benchmark* serão executados em ciclos de forma a colher os tempos médios para evitar distorções de picos ou ruídos na arquitetura.

6. ATIVIDADES JÁ REALIZADAS

Um levantamento bibliográfico preliminar foi realizado e resultou em diversos posicionamentos nos cenários de BI e RTBI em ambiente tradicional, ambiente em nuvem e BI em nuvem. A proposta inicial da solução foi especificada e as atividades para sua implementação foram iniciadas. O experimento foi iniciado com resultados preliminares promissores, que tendem a relativizar a perda de desempenho na nuvem, em relação às vantagens da solução proposta.

7. CONCLUSÃO

O panorama de BI em Tempo Real requer grande poder de processamento, armazenamento de dados e baixa latência na atualização dos dados da base de dados operacional para o data warehouse. Por outro lado, o ambiente em nuvem propõe uma "infinidade" de recursos a um custo aceitável, onde as empresas estão aportando seus recursos e migrando suas bases de dados para tal contexto. Com este estudo espera-se contribuir com uma definição de um ambiente e implementação de RTBI em nuvem para base de dados relacional, que proporcione um desempenho tolerável, ou seja, que as consultas não tenham sua performance acentuadamente prejudicadas quando as atualizações em tempo real ocorrerem.

8. REFERÊNCIAS

[1] Al-Aqrabi, H., Liu, L., Hill, R., Antonopoulos, N. (2014). Cloud BI: Future of business intelligence in the Cloud. Journal of Computer and System Sciences.

[2] Amazon, 2015, Disponível em: <http://aws.amazon.com/pt/ec2/> . Acesso em 14 Fev 2015

[3] Arzoumanian, J., Mustafa, S., 2014, Pratices that organizations employ to enhance business intelligence agility, Department of Informatics, Lund University, Sweden.

[4] Awoyelu, I. O., Udo, I. J., Omodunbi, T., (2014) Bridging the Gap in Modern Computing Infrastructures: Issues and Challenges of Data Warehousing and Cloud Computing. Computer and Information Science; Vol. 7, No. 1.

[5] F. Dehne, et al., Scalable real-time OLAP on cloud architectures, J. Parallel Distrib. Comput. (2014), http://dx.doi.org/10/1016/j.jpdc.2014.08.006

[6] Ferreira, N., Martins, P., Furtado, P., 2013, Near real-time with traditional data warehouse architectures: factors and how-to, Proceedings of the 17th International Database Engineering, Barcelona, Spain.

[7] Inmon, H. W., Strauss, D., Neushloss, G., (2008) DW 2.0: The Architecture for the Next Generation of Data Warehousing, Morgan Kaufmann.

[8] Jula A., Sundararajan E., and Othman Z.. 2014. Review: Cloud computing service composition: A systematic literature review. *Expert Syst. Appl.* 41, 8 (June 2014).

[9] Kabiri, A., Chiadmi D, (2014) Architecture for Near Zero Latency in Datawarehouse, Journal of Software, Vol 9, Nº 5.

[10] Muntean, M., Surcel, T. (2013). Agile BI - The Future of BI. Informatica Economica vol. 17.

[11] Kimball R., Ross M., (2013) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition, Wiley.

[12] Pereira, D., Azevedo, L. G., Tanaka, A., Baião, F. (2012). Real time data loading and OLAP queries: Living together in next generation BI environments. Journal of Information and Data Management, 3(2), 110.

[13] REAL-TIME-SSB, 2011, A. Real Time SSB Project. Disponível em <http://code.google.com/p/real-time-ssb/>, Acesso em : 20 Dez 2014.

[14] Schad, J., Dittrich, J. and Quiané-Ruiz, J., (2010), Runtime measurements in the cloud: observing, analyzing, and reducing variance, Proc. VLDB Endow. 3 (2010) 460-741.

[15] Verma, H.. 2013 "Data-warehousing on Cloud Computing." *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 2.2 (2013).

[16] Waas F.,Wrembel F, Freudenreich T., Thiele M. ,Koncilia C., and Furtado P.,(2013) "On-Demand ELT Architecture for Right-Time BI: Extending the Vision", ;presented at IJDWM, pp.21-38.

VMTools-RA - Uma Proposta de Arquitetura de Referência para Ferramentas de Gerenciamento de Variabilidades

Alternative Title: VMTools-RA - A Proposal of Reference Architecture for Variability Management Tools

Ana Paula Allian
Programa de Pós-Graduação em Ciência da Computação
Universidade Estadual de Maringá - UEM
Maringá, Paraná
ana.allian@gmail.com

Edson OliveiraJr
Programa de Pós-Graduação em Ciência da Computação
Universidade Estadual de Maringá - UEM
Maringá, Paraná
edson@din.uem.br

Elisa Nakagawa
Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional
Universidade de São Paulo - USP
São Carlos, São Paulo
elisa@icmc.usp.br

RESUMO

Gerenciar diferenças entre os produtos de software é uma das atividades essenciais para o sucesso do reúso de software. Tal atividade é conhecida por gerenciamento de variabilidade (GV) e é apoiada por diversas ferramentas. Em outro contexto, arquiteturas de referência (AR), arquiteturas genéricas para uma classe de sistemas, podem contribuir para o reúso de conhecimento sobre GV e padronização de ferramentas para tal domínio. Este trabalho apresenta a proposta da VMTools-RA, uma AR para ferramentas de GV. Espera-se que essa AR possa contribuir para o desenvolvimento de novas ferramentas de GV, além de fornecer suporte à integração e reúso de experiências em projetos.

Palavras-Chave

Arquitetura de Referência, Gerenciamento de Variabilidade.

ABSTRACT

Managing differences of software products is one of the essential activities to the success of software reuse. Such an activity is known as variability management (VM), supported by many tools. In another context, reference architectures (RA), generic architectures for a class of systems, can contribute to the reuse of knowledge on VM and standardization of tools for this domain. This paper presents VMTools-RA, a RA for VM tools. We intend this RA can contribute to the development of new VM tools, as well as providing support to integration and reuse of projects experiences.

SBSI 2015, May 26th-29th, 2015, Goiânia, Goiás, Brazil

Categories and Subject Descriptors

D.2.11 [**Software Engineering**]: Software Architecture

General Terms

Design, Standardization, Theory

Keywords

Reference Architecture, Variability Management.

1. INTRODUÇÃO

Variabilidade é o termo utilizado para representar como produtos podem se diferenciar entre si. Tal termo surgiu a partir da concepção de *frameworks* e da necessidade desses em serem customizados por meio de seus *hot spots* que são os pontos de variação[7]. Tal termo se consolidou em diferentes abordagens, em especial a de linha de produto de software [5, 3]. A identificação da variabilidade é feita através dos requisitos de sistema e por meio do modelo de *feature*, uma característica de um sistema que é relevante e visível para o usuário final, e descreve o domínio em que o sistema está inserido. Assim, o Gerenciamento de Variabilidades (GV) se tornou uma atividade imprescindível para o sucesso de tais abordagens.

Atualmente, existem diversas ferramentas de GV acadêmicas e industriais. Tais ferramentas apoiam desde a modelagem do domínio e das variabilidades até a geração de produtos específicos. Devido à heterogeneidade de ferramentas, a indústria tem utilizado diferentes tipos de soluções para gerenciar a variabilidade, além de criar suas próprias ferramentas [2], potencialmente reduzindo a transferência de tecnologia da academia, que é vital para qualquer campo de pesquisa. De acordo com o estudo de Berget et al.[2] essa variedade de notações e ferramentas de GV utilizadas na indústria ameaçam a abordagem de gerenciamento de variabilidade que se concentra em exatamente uma representação da variabilidade. Como consequência, um aumento na complexidade da atividade de GV é observada, exigindo atenção para a qualidade dessas ferramentas.

Apesar da diversidade de ferramentas de GV existentes, uma limitação comum a maioria das soluções disponíveis

refere-se a falta de padronização quanto as funcionalidades básicas dessas ferramentas. A identificação de funcionalidades essenciais para ferramentas de GV pode contribuir para o desenvolvimento de novas ferramentas de GV, além de fornecer suporte à integração e reúso de experiências em projetos.

Portanto, entende-se como uma oportunidade de pesquisa responder a seguinte questão: ***É possível propor uma padronização em nível arquitetural para ferramentas de gerenciamento de variabilidade?***. Para responder a tal questão pode-se explorar os conceitos relacionados a arquiteturas de referência.

Arquitetura de Referência (AR) é um tipo especial de arquitetura de software [6]. É considerada um padrão prédefinido projetado para um determinado contexto de negócio. Tal conceito faz uso de artefatos muitas vezes concebidos em projetos anteriores. Para isso, uma AR deve possuir conhecimento, experiência e informação adquirida. Além disso, uma AR alcança a compreensão de domínios específicos bem reconhecidos, promovendo a reutilização de experiências em projetos e facilitando o desenvolvimento, padronização, qualidade e evolução de sistemas de software [1, 6]

2. APRESENTAÇÃO DO PROBLEMA

O problema tratado neste trabalho é a heterogeneidade de GV e a dificuldade das empresas em estabelecer a prática do apoio automatizado para tal atividade.

O estudo de Berger et al.[2] aponta a carência de padronização de ferramentas de GV e o estado da prática na forma de experiências isoladas pela indústria. A proposta de uma padronização, mesmo em nível arquitetural, possibilitaria o desenvolvimento de ferramentas de GV, aumentando a interoperabilidade e permitindo guiar-se por soluções conhecidas e experimentadas pela academia e indústria. Além disso, há contribuição com o reúso e simplificação na tomada de decisão no desenvolvimento de novas ferramentas, o que se torna possível frente ao conhecimento estabelecido e documentado sobre GV. Para tanto, é possível explorar os conceitos que permeiam arquiteturas de referência.

Acredita-se, portanto, que o conceito de AR possa ser adotado como uma estratégia de se estabelecer uma padronização arquitetural para ferramentas de GV.

3. PROPOSTA DE SOLUÇÃO

Esta pesquisa tem como objetivo propor uma AR para ferramentas de GV seguindo o processo PROSA-RA. PROSA-RA é um processo iterativo que sistematiza as etapas para construir, representar e avaliar AR [6]. Esse processo é apoiado pelo metamodelo RAModel. Tal metamodelo representa os elementos essenciais, os quais podem estar contidos em uma AR. O PROSA-RA possui quatro etapas conforme pode ser visualizado na Figura 1:

Etapa 1. Investigação das fontes de informação: obtidas de pessoas, sistemas de software, publicações em periódicos, livros, outros modelos de referência, ontologias, entre outros. Para o desenvolvimento dessa etapa um estudo secundário na forma de uma Revisão Sistemática da Literatura (RSL) foi realizado[1]. Nesse cenário, para garantir a qualidade do protocolo da RSL, um conjunto de diretrizes estabelecido por Kitchenham [4] foi aplicado rigorosamente. Foram encontradas 43 ferramentas de GV, as quais são analisadas sob a ótica de vários critérios como as suas principais características, padrões de arquitetura, funcionalidades, abordagens de modelagem de características, maturidade quanto ao ciclo de vida de desenvolvimento e tecnologias de desenvolvimento. Somados a essa fonte de pesquisa, outras RSL sobre ferramentas de GV e estudos primários também foram empregados nessa etapa. Nesse contexto, foram considerados, a princípio, os seguintes grupos de informação: (i) Arquiteturas concretas e ferramentas de GV, (ii) Tecnologia de desenvolvimento para ferramentas de GV; (iii) Padronizações e conceitos aplicáveis ao contexto do domínio de GV; (iv) Arquiteturas de referências de outros domínios para auxiliar na construção dessa proposta (ainda está sendo avaliado qual AR).

Etapa 2. Estabelecimento dos requisitos arquiteturais: com base nas fontes de informação são identificados os requisitos de sistemas de software de domínio, de AR e conceitos que devem ser considerados nessa arquitetura, por exemplo, estilos arquiteturais, linguagem de programação, funcionalidades, atributos de maturidade, entre outros serão, então, mapeados aos requisitos arquiteturais que melhor se endereçam. Para apoiar a Etapa 2 do processo PROSA-RA será utilizado o metamodelo RAModel. Tal metamodelo estabelece quatro grupos de elementos que podem estar eventualmente contidos em uma AR: domínio, aplicação, infraestrutura e elementos transversais. O RAModel fornece um conjunto abrangente de elementos com a intenção de projetar arquiteturas completas. Como resultado, essa etapa visa elencar os requisitos arquiteturais de ferramentas de GV, por exemplo: (i) A AR deve possibilitar a representação das características variáveis e obrigatórias; (ii)A AR deve possibilitar a configuração do modelo de característica; (iii) A AR deve possibilitar a derivação de produtos; Além de outros requisitos que estão em fase de análise.

Etapa 3. Projeto arquitetural: é realizado utilizando os estilos e os padrões de arquitetura, bem como uma combinação desses e de outros estilos, provavelmente identificadas na Etapa 1. Os requisitos arquiteturais identificados serão utilizados como base para realizar o projeto arquitetural da **Etapa 3**. Descrições textuais e diagramas, fluxogramas ou UML por exemplo, podem ser utilizados para representar a AR em diferentes visões arquiteturais como estrutural, de implantação e de tempo de execução. A Figura 2 apresenta uma proposta inicial com alguns dos elementos já identificados que devem estar contidos no VMTools-RA. É importante observar que ainda falta as ligações desses elementos bem como uma estruturação mais formal dessa AR que está em fase de construção.

Etapa 4. Avaliação Arquitetural: verifica a descrição da arquitetura junto as partes interessadas com o objetivo de detectar defeitos na descrição da arquitetura. Tal etapa tem como objetivo avaliar a viabilidade e qualidade da AR.

4. AVALIAÇÃO DA SOLUÇÃO PROPOSTA

O objetivo principal durante a avaliação da AR é obter informações para averiguar: (i) se a AR está adequadamente representada; (ii) se a documentação relacionada à AR contém informações pertinentes; (iii) se são considerados os atributos de qualidade para a AR, tais como, a intero-

[1]Um artigo sobre a RSL está em fase de submissão ao periódico *Information and Software Technology*. sobre ferramentas de GV

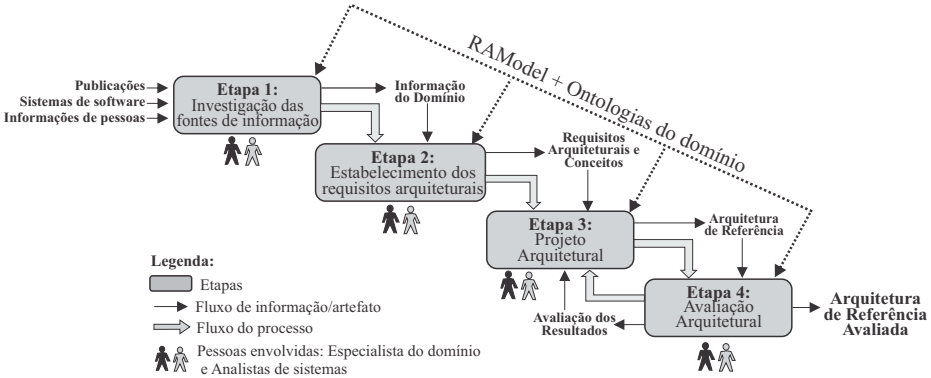

Figura 1: Etapas do Processo Prosa-RA. Traduzido de [6].

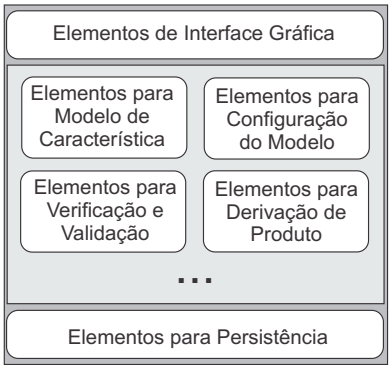

Figura 2: VMTools-RA - Proposta Inicial (Piloto)

perabilidade, segurança e escalabilidade; (iv) se a AR pode ser plenamente instanciada; e (v) o que pode ser alterado, se necessário, a fim de evoluir a AR.

Para a avaliação desta pesquisa será realizado um estudo empírico e um estudo de caso. O estudo empírico envolverá um estudo qualitativo, visando analisar a viabilidade da AR com base no conhecimento de especialistas. Tal estudo será apoiado por atividades de *Grounded Theory* como, por exemplo, *Coding* buscando codificar o conhecimento dos especialistas sobre a AR proposta. Além de entrevistas informais realizadas com especialistas de GV um questionário será elaborado, a fim de identificar defeitos relacionados à omissão, ambiguidade, inconsistência e de informações incorretas na VMTools-RA.

Já o estudo de caso envolverá a implementação de uma ferramenta de GV a partir da VMTools-RA. Para o desenvolvimento do sistema proposto, os requisitos arquiteturais instanciados da VMTools-RA serão elicitados. O objetivo desse estudo de caso é obter evidências da facilidade de uso da arquitetura de referência proposta.

5. ATIVIDADES REALIZADAS

Até o momento, foi realizada uma revisão sistemática da literatura (RSL) com o objetivo de congregar as ferramentas de GV existentes na acadêmia e na indústria, bem como suas principais características, padrões de arquitetura e funcionalidades. Além disso, prossegue-se com a ênfase no estudo da literatura com o objetivo de identificar os requisitos essenciais sobre tais ferramentas.

6. CONCLUSÃO

Os principais objetivos da proposta apresentada neste artigo estão relacionados a propor e avaliar uma AR para ferramentas de GV. Logo, uma RSL foi realizada, na qual foi possível identificar 43 estudos sobre ferramentas, bem como suas tecnologias, padrões de arquiteturas e maturidade da ferramenta relacionado ao ciclo de vida de desenvolvimento. Tais estudos auxiliarão no cumprimento dos objetivos definidos, no intuito de propor a AR, como também no planejamento e condução de estudos empíricos objetivando a avaliação da AR proposta.

7. REFERÊNCIAS

[1] S. Angelov, P. Grefen, and D. Greefhorst. A classification of software reference architectures: Analyzing their success and effectiveness. In *WICSA/ECSA, Cambridge, UK*, pages 141–150, 2009.

[2] T. Berger, R. Rublack, D. Nair, J. M. Atlee, M. Becker, K. Czarnecki, and A. Wsowski. A survey of variability modeling in industrial practice. In *VaMoS, Pisa, Italy*, pages 7:1–7:8, Jan 2013.

[3] R. Capilla, J. Bosch, and K. C. Kang. *Systems and software variability management, concepts, tools and experiences*, volume 14. Springer, 2013.

[4] B. Kitchenham, B. O. Pearl, D. Budgen, M. Turner, J. Bailey, and S. Linkman. Systematic literature reviews in software engineering. *Information and Software Technology*, 51(1):7–15, 2009.

[5] F. J. v. d. Linden, K. Schmid, and E. Rommes. *Software product lines in action: The best industrial practice in product line engineering*. Springer-Verlag, New York, 2007.

[6] E. Y. Nakagawa, M. Guessi, J. C. Maldonado, D. Feitosa, and F. Oquendo. Consolidating a process for the design, representation, and evaluation of reference architectures. In *WICSA, Sydney, NSW*, pages 143–152, Apr 2014.

[7] W. Pree and H. Sikora. Design patterns for object-oriented software development (tutorial). In *Proceedings of the 19th Int. Conf. on ICSE*, pages 663–664, 1997.

Índice Remissivo