

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

Raphael A. Rodrigues¹, Leonardo G. Azevedo^{1,2}, Kate Revoredo¹, Henrik Leopold³

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

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

³WU Vienna, Welthandelsplatz 1, 1020 Vienna, Austria

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

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

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

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

1. Introdução

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

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

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

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

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

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

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

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

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

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

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

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

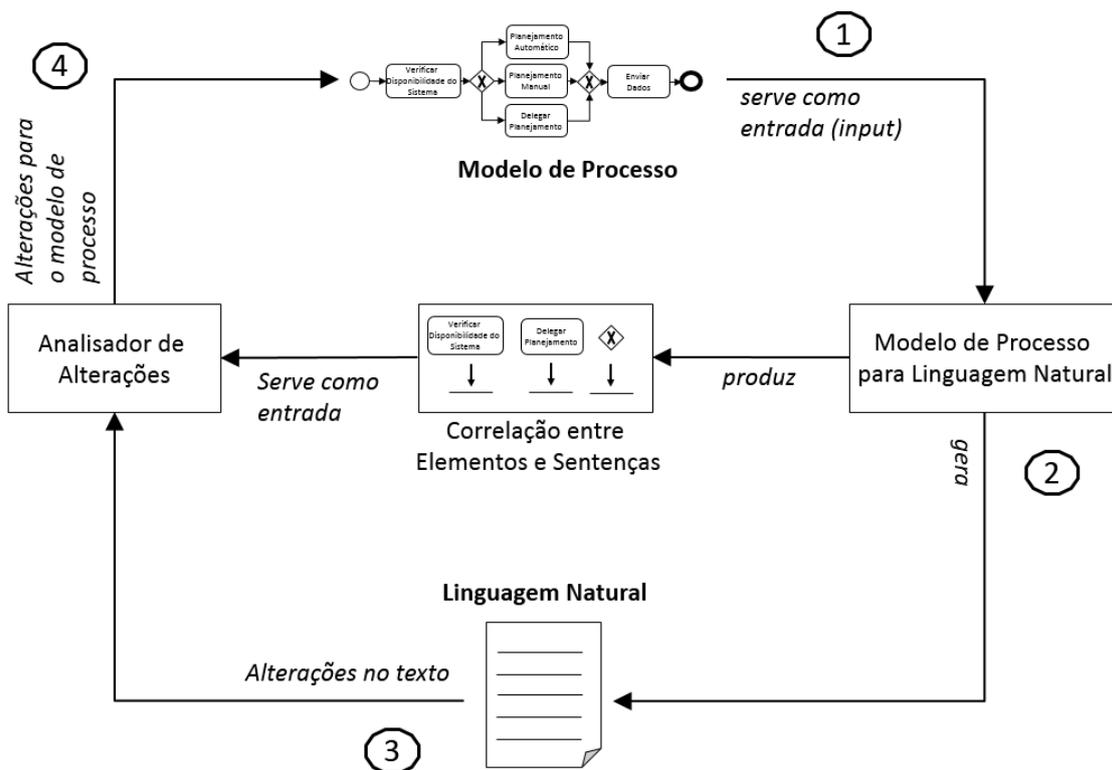


Figura 1. Exemplo da arquitetura para a solução proposta.

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

2. Proposta de Projeto de Pesquisa

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

2.1. Visão Geral da Proposta

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

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

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

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

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

2.2. Objetivos da Proposta

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

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

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

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

2.3. Metodologia

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

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

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

3. Conclusão

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

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

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

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

Referências

- Dalianis, H. (1992). A method for validating a conceptual model by natural language discourse generation. In *Advanced Information Systems Engineering*, pages 425–444. Springer.
- Davies, I., Green, P., Rosemann, M., Indulska, M., and Gallo, S. (2006). How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3):358–380.
- Figl, K. and Laue, R. (2011). Cognitive complexity in business process modeling. In *Advanced Information Systems Engineering*, pages 452–466. Springer.
- Kautz, K., Hansen, B., and Jacobsen, D. (2004). The utilization of information systems development methodologies in practice.
- Kettinger, W. J., Teng, J. T., and Guha, S. (1997). Business process change: a study of methodologies, techniques, and tools. *MIS quarterly*, pages 55–80.
- Ko, R. K., Lee, S. S., and Lee, E. W. (2009). Business process management (bpm) standards: a survey. *Business Process Management Journal*, 15(5):744–791.
- Lavoie, B., Rambow, O., and Reiter, E. (1996). The model explainer. In *Demonstration presented at the Eighth International Workshop on Natural Language Generation, Herstmonceux, Sussex*.

- Leão, F., Revoredo, K., and Baião, F. (2013). Learning well-founded ontologies through word sense disambiguation. In *Intelligent Systems (BRACIS), 2013 Brazilian Conference on*, pages 195–200. BRACIS.
- Leopold, H., Mendling, J., and Polyvyanyy, A. (2012). Generating natural language texts from business process models. In *Advanced Information Systems Engineering*, pages 64–79. Springer.
- Leopold, H., Mendling, J., and Polyvyanyy, A. (2014). Supporting process model validation through natural language generation. In *Transactions on Software Engineering* 40(8), pages 818–840. IEEE.
- Mendling, J., Strembeck, M., and Recker, J. (2012). Factors of process model comprehension findings from a series of experiments. *Decision Support Systems*, 53(1):195–206.
- Meziane, F., Athanasakis, N., and Ananiadou, S. (2008). Generating natural language specifications from uml class diagrams. *Requirements Engineering*, 13(1):1–18.
- Nijssen, G. M. and Halpin, T. A. (1989). *Conceptual Schema and Relational Database Design: a fact oriented approach*. Prentice-Hall, Inc.
- OMG (2011). Business process model and notation (bpmn) version 2.0. <http://www.bpmn.org/>.
- Reijers, H. A. and Mendling, J. (2011). A study into the factors that influence the understandability of business process models. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(3):449–462.
- Rodrigues, R. (2013). *Um Framework Genérico para Geração de Texto em Linguagem Natural a partir de Modelos de Processo de Negócio*. Bachelor thesis in Information Systems – Federal University of the State of Rio de Janeiro (UNIRIO).