

Uma Abordagem para Gerar Texto Orientado a Processo a partir de Texto em Linguagem Natural

Alternative Title: An Approach to Generate Process-oriented Text from Natural Language

Renato César Borges Ferreira
Programa de Pós-Graduação em Computação
Instituto de Informática
Universidade Federal do Rio Grande do Sul
Porto Alegre, Brasil
renato.borges@inf.ufrgs.br

Lucinéia Heloisa Thom
Programa de Pós-Graduação em Computação
Instituto de Informática
Universidade Federal do Rio Grande do Sul
Porto Alegre, Brasil
lucineia@inf.ufrgs.br

RESUMO

Em organizações, a modelagem de processos de negócio é de grande importância para documentar, entender e automatizar processos. Essas organizações geralmente provêm documentos não estruturados e de difícil entendimento por parte dos analistas. A extração de modelos ou fragmentos de processo a partir de descrições textuais pode contribuir para minimizar o esforço necessário à modelagem de processos. Neste contexto, o presente artigo propõe uma abordagem que visa gerar texto orientado a processo a partir de texto em Linguagem Natural. Este artigo propõe investigar a estrutura que um texto em Linguagem Natural deve apresentar para que a partir deste se possa extrair fragmentos ou modelos de processo. Com base no estudo sobre classes gramaticais realizado no contexto desta pesquisa, não existe uma ordem estabelecida ou padronizada em que as classes gramaticais devem ser apresentadas no texto em Linguagem Natural. Desta forma, resultados preliminares desta pesquisa mostram que através de regras de mapeamento e correlações entre palavras representam que as classes gramaticais indicam um elemento de processo, através de palavras chave e/ou tempos verbais.

Palavras-Chave

Texto Orientado a Processo, Linguagem Natural, Gerenciamento de Processo de Negócio, Notação de Modelagem de Processos de Negócio.

ABSTRACT

In organizations, the business process modeling is of big importance to report, understand and automate processes. These organizations usually come unstructured documents and difficult to understand by analysts. The extraction models or fragments process from textual descriptions may contribute to minimizing the effort required to process modeling. In this context, this paper proposes an approach to

generate text-oriented process from the text in natural language. This paper proposes to investigate the structure of a text in natural language must show to from this one can extract fragments or process models. Based on the study of grammar classes accomplished in the context of this research, there is no established or standardized order in which the grammar classes should be shown in the text in natural language. Thus, preliminary results of this research show that by mapping rules and correlations between words represent the grammatical classes indicate a process element, through keywords and/or verb tenses.

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]: Natural Language; H.4 [Information Systems]: Information Systems Applications; D.2 [Software]: Software Engineering

General Terms

Mapping Rules, Correlation, Labels, Sequential Texts, Process Models.

Keywords

Oriented Process text, Natural Language, Business Process Management, Business Process Model and Notation.

1. INTRODUÇÃO

Organizações públicas ou privadas para manterem-se competitivas no mercado, buscam melhor interação com clientes e parceiros de negócio, ofertar um melhor produto ou serviço para seus clientes e atingir uma melhor padronização e eficiência na execução de seus processos. A automatização de processos executados em uma organização, proporciona maior controle sobre custos, tempo, erros e redundância na execução dos processos [11][10].

Um processo de negócio consiste de uma série de atividades relacionadas, que são realizadas para atingir um determinado objetivo de negócio [1]. Conforme [7], um processo inclui uma sequência ou fluxo de atividades que são executados com o objetivo de realizar um trabalho. Segundo [2], processo de negócio é uma coleção de eventos, atividades e tomadas de decisões inter-relacionadas, as quais envolvem uma série de atores e objetos, que coletivamente levam a resultados que agregam valor para pelo menos um cliente. Um exemplo de processo de negócio seria um processo de compra e venda de um produto, onde existe uma tarefa de verificação da ordem de compra, um evento de envio do produto e uma próxima tarefa de recebimento de produto.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SBSI 2016, May 17th-20th, 2016, Florianópolis, Santa Catarina, Brazil
Copyright SBC 2016.

O Gerenciamento de Processo de Negócio (BPM – *Business Process Management*) pode ser definido como um conjunto de métodos, técnicas e ferramentas para descoberta, análise, redesenho, execução e monitoramento de processos de negócio [12] [2]. Segundo [4], através de BPM organizações podem se adaptar com flexibilidade em um ambiente de negócios em constante mudança. Portanto, BPM traz inúmeras melhorias para a organização, tal como a padronização de processos, melhoria, qualidade e rapidez na execução das atividades [11] [10].

O ciclo de vida de BPM inclui 6 etapas: (i) identificação de processos, (ii) descobrimento, (iii) análise, (iv) redesenho, (v) automatização e (vi) monitoramento e controle de processos [2]. Cada etapa do ciclo de vida é de extrema importância para automatização de processos dentro da organização.

A modelagem de processo pode ser considerada uma das mais importantes e duradouras etapas do ciclo de vida de BPM. Tal etapa tem como finalidade facilitar a concepção do processo e associar o conhecimento sobre o processo entre os envolvidos em sua execução, assim como identificar e prever possíveis problemas organizacionais. Em [2], é relatado que a modelagem de processos é pré-requisito para a análise, redesenho e automação de processos de negócio. Uma modelagem de processos incorreta compromete as próximas etapas de BPM, ou seja, uma correta automatização de processos origina-se de uma precisa modelagem de processos.

Nós aprendemos com a prática que a concepção de processos particulares (ex.: processos da saúde) podem ser complexos, não somente pela sua variedade e necessidade de flexibilidade, mas também porque requerem o conhecimento de diversos termos do domínio, que podem levar a ambiguidades e problemas de interpretação entre analistas de processos e especialistas de domínio. Em vista disso, na fase de modelagem de processos, existem diversos métodos de identificação de processos, tais como: entrevistas com usuários, *workshops* sobre o tema e documentos existentes na organização para iniciar a modelagem dos processos [9]. Porém, tais métodos podem apresentar limitações devido a problemas de comunicação entre analistas e usuários, falta de documentação e padronização dos processos e ausência de informações de usuários.

Em vista disso, neste artigo propomos uma abordagem para gerar texto orientado a processo a partir de texto em Linguagem Natural. Observamos que a maioria dos trabalhos descritos na literatura, consideram que os textos em Linguagem Natural, estão gramaticalmente corretos para geração de modelos de processos de negócio ou vice-versa. Uma das principais motivações para este artigo é investigar a estrutura que um texto em Linguagem Natural deve apresentar para que, a partir deste se possa extrair fragmentos ou modelos de processo. Ou seja, a hipótese deste artigo é que é possível gerar texto orientado a processo a partir de texto em Linguagem Natural.

Neste artigo, texto orientado a processo é aquele que apresenta uma estrutura que possibilita identificar papel associado a uma atividade, piscinas relacionadas com raias, interação entre piscinas (fluxo de mensagem), identificar eventos (início, intermediário e fim) e apontar pontos de decisão (*AND*, *OR* e *XOR*). Espera-se gerar um modelo (*template*) de como o texto deve ser estruturado para extração de modelos de processo a partir de texto.

A extração de modelos de processo a partir de texto pode minimizar o esforço por parte do analista de processos para capturar, principalmente através de entrevistas com usuários os processos de negócio executados na organização. Nem sempre estas entrevistas levam ao conhecimento global do processo, ou seja, tais entrevistas possibilitam o entendimento, muitas vezes, apenas de parte isoladas do processo.

Na Língua Inglesa existem classificações de textos. Cada

classificação apresenta características diferentes, onde essas características são palavras, frases e questões relacionadas com cada particularidades do texto. Exemplos de classificações incluem: Textos Descritivos; Comparativos e contraste; Ordem de importância; Problema e solução; Causa e efeito; Sequenciais. Em textos sequenciais, a informação é organizada em etapas ou um processo é explicado na ordem em que ocorre. [13]

As características presentes no texto de classificação sequencial tem particularidades presentes em modelos de processo de negócio. Existem palavras chave que são correntes nos textos sequenciais, tais como: *primeiro*, *segundo*, *próximo*, *então*, *finalmente*, *seguinte*, *agora*, *depois*, entre outras. Tais palavras apresentam possíveis relações (correlações) com elementos de modelagem da Notação e Modelo de Processos de Negócio (BPMN¹ – *Business Process Model and Notation* (ex.: atividades, *gateways*, piscinas, raias, etc). Diante desses aspectos, a abordagem deste artigo atribuirá como referência o texto de classificação sequencial.

O restante deste artigo está estruturado do seguinte modo: A seção 2 apresenta trabalhos relacionados de extração de modelos de processo a partir de textos e vice-versa. A seção 3 mostra a abordagem proposta de geração de texto orientado a processo a partir de Linguagem Natural. A seção 4 relata as discussões de resultados preliminares da abordagem. E por fim, a seção 5 apresenta as conclusões deste artigo.

2. TRABALHOS RELACIONADOS

Trabalhos acadêmicos estão sendo realizados sobre extração de texto a partir de modelos de processo ou geração de modelos de processo a partir de texto. Em [3], é proposta a extração de modelos de processo de negócio a partir de descrições textuais. A abordagem proposta considera três aspectos: análise sintática, determinação de uma árvore de sintaxe e relações gramaticais entre as partes da sentença; análise semântica, extração do significado de palavras ou frases; resolução de anáfora, identificação de conceitos que são referências usando pronomes (*we*, *he* e *it*) e artigos (*this*, *that*).

As limitações desta abordagem, partem do pressuposto que os textos analisados para a geração de modelos de processo, estão gramaticalmente corretos na Língua Inglesa, isto é, retira-se e/ou corrige manualmente palavras ou frases que estão incorretas gramaticalmente, tornando o texto a ser analisado pela ferramenta, gramaticalmente correto.

Em [4] é apresentado uma abordagem para extração de textos em linguagens natural a partir de modelos de processo. Nessa abordagem, o autor retrata desafios para obter textos a partir de modelos de processo: Planejamento do texto; Planejamento de sentença; Realização de superfície e Flexibilidade. Para a proposta apresentada neste artigo é de especial interesse a abordagem do planejamento de texto. Essa abordagem esta dividida em três etapas: Extração linguística de Informação, onde ocorre a extração de componentes linguísticos dos *labels* dos elementos dos modelos de processo; Geração *Annotated rpST*, faz a linearização do modelo de processo através da geração de uma estrutura de árvore; Estruturação do texto, onde ocorre a aplicação de técnicas de estrutura de texto, como a inserção de *n* pontos com base na estrutura de árvore computadorizada.

As limitações presentes neste trabalho mostram evidências, em que as frases geradas pela ferramenta são comparativamente curtas e simples. Outra limitação, ocorre em garantir um nível estável de complexidade entre os textos criados manualmente, assim, seria necessário treinar os escritores de texto. Por fim, se a modelagem de processo não

¹BPMN é uma notação padrão para modelagem de processos adotada pela *Object Management Group (OMG)* e é usada por ferramentas de modelagem (ex.: Bizagi, Intalio, etc).

está bem documentada, o texto gerado conseqüentemente não estará estruturado.

3. GERAÇÃO DE TEXTO ORIENTADO A PROCESSO A PARTIR DE TEXTO EM LINGUAGEM NATURAL

Neste artigo, propomos uma abordagem para gerar texto orientado a processo a partir de texto em Linguagem Natural. A Figura 1, ilustra a abordagem proposta. Tal abordagem deve auxiliar no processo de modelagem de processos e apoiar o trabalho do analista nesta fase do ciclo de vida de BPM.

A Figura 1 mostra o texto em Linguagem Natural, incluindo sentenças não estruturadas, que não correspondem a textos sequenciais e conseqüentemente dificultam a extração de modelos de processo a partir dele. Sendo assim, foram elaboradas regras de mapeamento e correlações de palavras para identificar possíveis modelos ou fragmentos de processo objetivando obter o texto orientado a processo.

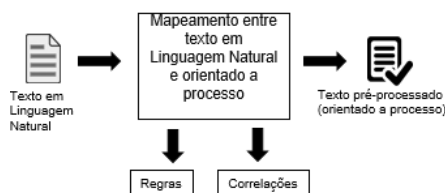


Figura 1: Abordagem de geração de texto orientado a processo a partir de texto em Linguagem Natural.

As regras se originam de um diversificado conjunto de classes gramaticais (*verbo, pronome, artigo, numeral e substantivo*). Com base no estudo sobre classes gramaticais realizado no contexto desta pesquisa, não existe um padrão em que as classes gramaticais devem ser apresentadas no texto. Sendo assim, elas estão relacionadas entre si e representam um elemento de processo. Por exemplo, as sentenças que contêm *Sujeito, Verbo e Objeto* representam um elemento de processo, tal como: Tarefa Manual e não dependem da ordem em que aparecem no texto. Em algumas sentenças, as classes gramaticais mesclam entre si, por exemplo: sentenças que contêm a ordem de *Verbo*, seguido de um *Sujeito* e posteriormente um *Objeto*, representam uma tarefa manual do processo e estão posicionadas de maneiras diferentes na sentença.

As regras foram definidas manualmente, sendo que cada regra é classificada conforme uma categoria do conjunto de elementos reduzidos da BPMN (ex.: objetos de fluxo, conectores e divisões) proposto pela OMG (*Object Management Group*) e pelo fato de serem recorrentes em processos de negócio. Nesta fase da pesquisa, foram criadas regras para Atividades (tarefas manuais e serviços) e Eventos. Posteriormente, serão elaboradas regras para outros elementos contidos no conjunto reduzido da BPMN. Assim, as regras propostas neste artigo se relacionam as categorias de objetos de fluxo. As regras de objetos de fluxo, foram criadas para tarefas atômicas, ou seja, representam ações no processo que serão executadas por um participante ou sistema. Deste modo, as tarefas frequentemente terão um participante (*Sujeito* da sentença). A partir do *Sujeito* da sentença é definido o tipo de tarefa a ser identificada, tal como: Manual, onde o *Sujeito* é representado por um ser humano; Tarefa de serviço, onde o *Sujeito* é representado por um *software* ou equipamento eletrônico. Os rótulos (*labels*²) das tarefas partem do

²O uso dos *labels* é considerado uma boa prática de modelagem, pois impacta diretamente na clareza e compreensão

Tabela 1: Regras para identificação de Atividades (Tarefas).

ATIVIDADES			
Regra		Sentença	Exemplo de label
1	<Suj> + <Verb> + <Obj>	The Support Officer (Suj) updates (verb) all group calendars (obj).	Updates group calendars
2	<Obj> + <Suj> + <Verb>	The severity (Obj) of the claimant (Suj) is evaluated (Verb).	Evaluated severity
3	<Obj> + <Verb> + <Suj>	The forms (Obj) are send (Verb) to the claimant (Suj).	Send forms
4	<Verb> + <Artigo> + <Obj>	Choose (Verb) a (Art) document (Obj).	Choose document

pressuposto que na sentença, existirá no mínimo um *Sujeito, Verbo e Objeto*. Conforme [6] e [5], os *labels* das atividades são representados pelo *Verbo* e em seguida o *Objeto*. Por exemplo, na sentença, "O cliente envia o formulário para o analista", o *label* da atividade se tornaria: *Enviar Formulário* e o participante: *Cliente*. A Regra 2 (dois) da Tabela 1, mostra outro exemplo de identificação de atividades em sentenças. A regra contempla a sequência de *Objeto*, seguido de um *Sujeito* e posteriormente um *Verbo*, em que o *label* se tornaria a união do *Verbo e Objeto*.

As regras criadas para eventos, seguem o mesmo pressuposto das atividades, porém existem diferenças entre elas. As sentenças que representam tempos verbais no presente e/ou futuro representam Atividades. As sentenças que retratam tempos verbais no passado e/ou presente perfeito são classificados como Eventos, pois, conforme [7], Eventos são criados quando a sentença causa impacto ou ocorre algo instantaneamente, modificando o fluxo do processo. Outra diferença, ocorre na elaboração dos *labels* dos Eventos. Conforme [6] e [4], os *labels* de Eventos são representados através do *Objeto* da sentença e em seguida o *Verbo* no particípio. A Regra 3 (três) da Tabela 2, mostra um exemplo de identificação de Eventos em sentenças. A regra contempla a sequência de *Sujeito*, seguido de um *Objeto* e posteriormente um *Verbo*, em que o *label* da sentença se tornaria a união do *Objeto e Verbo*.

Tabela 2: Regras para identificação de Eventos.

EVENTOS			
Regra		Sentença	Exemplo de label
1	<Obj> + <Verb> + <Adj>	Once the forms (Obj) are returned, they are checked (Verb) for completeness (Adj).	Completeness checked
2	<Obj> + <Verb> + <Suj>	When the report (Obj) is received (Verb) by analyst (Suj).	Report received
3	<Suj> + <Obj> + <Verb>	The claimant (Suj) was informed that claim (Obj) must be rejected (Verb).	Claim Rejected
4	<Obj> + <Verb>	Urgent document (Obj) has been received (Verb).	Document Received

Com o objetivo de obter um texto orientado a processo e tendo as regras de mapeamento devidamente criadas, serão adotados as seguintes etapas para a abordagem de extração de modelos de processo a partir de texto: **1.** Marcar cada sentença do texto em Linguagem Natural: identificar sentenças/frases no texto e marcá-las. **2.** Selecionar as regras para serem aplicadas no texto marcado: identificação das regras criadas de acordo com tempo verbal e/ou palavras chave para serem empregadas no texto (ainda em Linguagem Natural). **3.** Aplicar as regras de mapeamento: aplicação da do processo. [5]

regra selecionada (etapa anterior) no texto que possibilite identificar o elemento de processo contido na sentença. 4. Gerar texto orientado a processo: Geração do texto orientado a processo (pré-processado) a partir do texto em Linguagem Natural com regras de mapeamento aplicadas.

Portanto, espera-se gerar um modelo (*template*) do texto orientado a processo para realizar a extração de fragmentos ou modelos de processo a partir de textos em Linguagem Natural.

3.1 Exemplo de Aplicação das Regras em Textos em Linguagem Natural

O Texto 1 em Linguagem Natural, retirado de um trabalho da área de Bioinformática [8], retrata a extração de elementos de processo a partir das regras criadas nesta abordagem.

Texto 1: ... Given a specific “template”, each fragment of a model is inserted into the template in order. The energy difference of *i*th fragment (*dE_i*) on dihedral angles of the template is simply measured by SimFold energy *E* ... To generate offspring, first, select a fragment using positive *dE* roulette ...

A sentença: “... each **fragment** of a model is **inserted** into the template in order ...” contempla a Regra 4 (quatro) de Eventos, incluindo *Verbo (Inserted)* no passado e o *Objeto (fragment)*. Ao realizar a extração do elemento de processo da sentença, o *label* do elemento se tornaria: *Fragment Inserted*.

Em seguida, ao realizar a análise da sentença: “... The energy difference of *i*th fragment (*dE_i*) on dihedral angles of the template is simply measured by SimFold energy *E* ...”, a mesma contempla a Regra 2 (dois) de Eventos, incluindo um *Sujeito (SimFold Energy E)*, um *Verbo (Measured)* no passado e o *Objeto (Energy)*. Ao realizar a extração do elemento de processo da sentença, o *label* e a raia do processo se tornaria, respectivamente: *Energy Measured* e *SimFold Energy E*.

Finalmente, ao realizar a análise da seguinte sentença: “... To generate offspring, first, **select a fragment** using **positive dE roulette** ...” é identificado uma atividade (tarefa de serviço). A sentença abrange a Regra 4 (quatro) de atividades, contendo um *Verbo (Select)*, seguido de um *Artigo (a)* e posteriormente um *Objeto (fragment)*. Ao realizar a extração do elemento de processo da sentença, o *label* e a raia do processo se tornaria, respectivamente: *Select Fragment* e *Positive dE Roulette*.

Com as regras de mapeamento aplicadas no texto, as sentenças no texto em Linguagem Natural representam um elemento de processo e consequente, geram um novo texto orientado a processo, isto é, identifica-se se mantém sentenças que representam elementos de processo e retira-se sentenças que não contemplam as regras de mapeamento criadas nesta abordagem.

4. DISCUSSÃO

As regras criadas inicialmente, mostram uma possível alternativa para realizar o mapeamento de texto em Linguagem Natural para texto orientado a processo. Nesta fase da pesquisa, foram elaboradas, apenas para 2 (dois) elementos de processo da BPMN, tais como: Atividades (tarefa manual e serviços) e Eventos, pois estes elementos são recorrentes em modelos de processo. Foi observado que as classes gramaticais das regras criadas, não apresentam ordem estabelecida ou padronizada, as mesmas mesclam entre si e indicam um elemento de processo, através de palavras chave e/ou tempos verbais.

Sobre o estudo dos tempos verbais da sentença, observamos que sentenças que contém tempos verbais no presente e/ou futuro, representam atividades. Sentenças em tempos verbais no passado e/ou presente perfeito da Língua Inglesa

representam Eventos.

5. CONCLUSÃO

A abordagem apresentada neste artigo mostra evidências que existe a possibilidade de gerar texto orientado a processo a partir de textos em Linguagem Natural.

Abordagens existentes, demonstram limitações sobre como os textos são preparados para realizar a extração de modelos/fragmentos de processo, evidenciando que retira-se e/ou corrige manualmente palavras ou sentenças que estão incorretas gramaticalmente. Deste modo, a abordagem apresentada neste artigo visa definir a estrutura que um texto em Linguagem Natural deve apresentar auxiliando na extração de modelos de processo.

Para abordagem proposta, investiga-se o desenvolvimento de um protótipo que possibilite, com base nas regras de mapeamento entre texto em Linguagem Natural e texto pré-processado a geração semiautomática de texto orientado a processo a partir de Linguagem Natural. Essa automatização poderá ser desenvolvida com técnicas e ferramentas de Processamento de Linguagem Natural (ex.: *Part-of-speech tagging*, *Stanford Parser*, *WordNet*, etc) para realizar a análise da sentença. Para validar a abordagem, serão criados formulários através do *Google Forms*, onde usuários (estudantes, desenvolvedores e professores) irão definir o texto para extração de modelos de processo com e sem o uso da abordagem proposta neste artigo.

6. REFERÊNCIAS

- [1] W. M. Coalition. Wfmc: Process definition language: Xpdl 2.0. page 164, 2005.
- [2] M. Dumas, M. L. Rosa, J. Mendling, and H. A. Reijers. *Fundamentals of Business Process Management*. Springer, 2013.
- [3] F. Friedrich, J. Mendling, and F. Puhmann. *Process model generation from natural language text*. 2011.
- [4] H. Leopold. *Natural language in business process models*. Springer, 2013.
- [5] J. Mendling, H. A. Reijers, and J. Recker. Activity labeling in process modeling: Empirical insights and recommendations. *Inf. Syst.*, 35(4):467–482, June 2010.
- [6] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst. Seven process modeling guidelines (7pmg). *Inf. Softw. Technol.*, 52(2):127–136, Feb. 2010.
- [7] OMG. Business process modeling notation (bpmn). versão 2.0.2, 2013.
- [8] S.-J. Park. A study of fragment-based protein structure prediction: biased fragment replacement for searching low-energy conformation. *Genome informatics. International Conference on Genome Informatics*, 2005.
- [9] I. Sommerville. *Software Engineering: 8th Edition, International Computer Science*. Addison-Wesley Longman Publishing, 2006.
- [10] L. Thom, M. Reichert, and C. Iochpe. Activity patterns in process-aware information systems: Basic concepts and empirical evidence. *International Journal of Business Process Integration and Management (IJBPIIM)*, 2009.
- [11] L. H. THOM. *Gerenciamento de Processos de Negócio e Aplicabilidade na Saúde e na Robótica*. Biblioteca Digital Brasileira de Computação, 2012.
- [12] M. Weske. *Business Process Management: Concepts, Languages, Architectures*. Springer-Verlag, Berlin, 2007.
- [13] R. Worksheets. Text structure. <http://www.ereadingworksheets.com/text-structure/>, page 1, 2014.