

Association for Information Systems AIS Electronic Library (AISeL)

Proceedings of the XI Brazilian Symposium on
Information Systems (SBSI 2015)

Brazilian Symposium on Information Systems
(SBIS)

5-2015

Ontology Extraction from Stories: an exploratory study in storytelling

Valdemar Tadeu Confort

Universidade Federal do Estado do Rio de Janeiro, valdemar.confort@uniriotec.br

Kate Revoredo

UNIRIO, katerevoredo@uniriotec.br

Fernanda Baião

Universidade Federal do Estado do Rio de Janeiro, Brazil, fernanda.baiao@uniriotec.br

Flavia Santoro

Federal University of the State of Rio de Janeiro, flavia.santoro@uniriotec.br

Follow this and additional works at: <http://aisel.aisnet.org/sbis2015>

Recommended Citation

Confort, Valdemar Tadeu; Revoredo, Kate; Baião, Fernanda; and Santoro, Flavia, "Ontology Extraction from Stories: an exploratory study in storytelling" (2015). *Proceedings of the XI Brazilian Symposium on Information Systems (SBSI 2015)*. 68.
<http://aisel.aisnet.org/sbis2015/68>

This material is brought to you by the Brazilian Symposium on Information Systems (SBIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the XI Brazilian Symposium on Information Systems (SBSI 2015) by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Extração de Ontologias a partir de Histórias: um estudo exploratório em *storytelling*

Alternative Title: Ontology Extraction from Stories: an exploratory study in *storytelling*

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

Kate Revoredo
Univ. Federal do Estado do Rio de
Janeiro
Av. Pasteur, 458, Urca
Rio de Janeiro, RJ, Brasil
katerevored@uniriotec.br

Fernanda Araujo Baião
Universidade Federal do Estado do
Rio de Janeiro
Av. Pasteur, 458, Urca
Rio de Janeiro, RJ, Brasil
fernanda.baiao@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

RESUMO

Empresas e sistemas de TI estão enfrentando um aumento na complexidade do ambiente caracterizada pela colaboração, mudança e variedade de clientes, fornecedores e produtos. Aplicar a técnica de *group storytelling* pode contribuir para a Gestão do Conhecimento da organização. A contagem de histórias traz benefícios desde a captura até fixação da informação, passando pela comunicação e entendimento dos conceitos. Empresas americanas (3M e Apple), japonesas (Sony e Toshiba) e europeias (ClubMed e Océ) já utilizam esta abordagem na prática. Por outro lado, a Engenharia de Ontologias pode contribuir na melhoria da qualidade da informação e oferecer uma solução para lidar com a gestão do conhecimento de forma sistemática. No entanto, a especificação e gestão de ontologias realizadas de forma manual podem ser caras, tediosas, enviesadas e propensas a erro. O aprendizado automático de ontologias é uma abordagem que extrai ontologias a partir de dados, tanto estruturados como não estruturados (textos). Este trabalho apresenta, em fase exploratória, uma proposta capaz de especificar automaticamente os elementos que compõem uma ontologia, a partir do conhecimento tácito dos envolvidos no domínio. Um estudo exploratório foi capaz de obter os conceitos de uma ontologia, de forma automática, a partir de histórias contadas numa ferramenta de *group storytelling* sobre os processos de negócio de uma das secretarias de uma Universidade de Ensino Superior Federal.

Palavras-Chave

Gestão do Conhecimento, Group Storytelling, Aprendizado de ontologias.

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

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

ABSTRACT

Business and IT systems are facing increasingly complex environments characterized by collaboration, change and variety of customers, suppliers and products. Group storytelling technique can contribute to the business knowledge management. The stories count brings benefits from capture to securing information, through communication and understanding of the concepts. American Companies (3M and Apple), Japanese (Sony and Toshiba) and European (ClubMed and Océ) already use this approach in practice. Ontology Engineering can contribute towards improving the quality of information and offer a solution to address knowledge management systematically. However, the specification and manually made of ontology management can be expensive, tedious, biased and prone to error. Automatic learning ontology is an approach that extracts ontology from the data, both structured and unstructured (text). This work presents, at the exploratory stage, a proposal able to specify, automatically, elements of an ontology, from the tacit knowledge of those involved in the field. An exploratory study was able to get the concepts of an ontology, automatically, from stories told by a group storytelling tool on the business process of one department of the University.

Categories and Subject Descriptors

I.2.6 [Computing Methodologies]: Artificial Intelligence – Learning *Knowledge Acquisition*

I.2.7. [Computing Methodologies]: Artificial Intelligence – Natural Language Processing *Text analysis*

H.4.1 [Information Systems]: Information Systems Applications – Office Automation *Groupware*

General Terms

Management, Documentation, Experimentation.

Keywords

Knowledge Management, Group storytelling, Ontology Learning.

1. INTRODUÇÃO

Empresas estão cada vez mais dependentes de suas redes de negócio, tendo que colaborar com outras organizações, lidar com uma maior variedade de produtos, clientes e infraestrutura organizacional [1] [2]. A Gestão do Conhecimento pode auxiliar as organizações a lidar com este desafio sob a perspectiva da gestão da informação e comunicação.

Registrar as informações relacionadas à execução de tarefas é essencial para a Gestão do Conhecimento. Esse registro pode ser feito através de histórias utilizando a técnica de *group storytelling* [2]. Porém, além da explicitação do conhecimento real através da técnica de *group storytelling*, é necessário manter a base de conhecimento sólida e com as ambiguidades semânticas controladas o máximo possível. Nesse contexto, a Engenharia de Ontologias pode contribuir na melhoria da qualidade dos modelos conceituais [3] que, por sua vez, contribuem na comunicação e troca de informação efetiva entre os participantes, sejam pessoas ou sistemas.

No entanto, a extração manual dos conceitos que comporão a ontologia pode ser cara, tediosa, propensa a erro, enviesada para o seu desenvolvedor, inflexível e específica para o propósito que motiva sua construção [4]. Aprendizado automático de ontologias a partir de dados não estruturados (texto) é uma alternativa capaz de lidar com as dificuldades apontadas.

Diante do cenário apresentado, a questão desta pesquisa é: Como especificar automaticamente uma ontologia a partir do conhecimento tácito dos envolvidos no domínio? A proposta de solução desta pesquisa é utilizar *group storytelling*, combinado com aprendizado de ontologias baseadas em mineração de texto. Desta forma, as histórias são construídas colaborativamente pelos envolvidos do domínio, e estas utilizadas para especificar automaticamente uma ontologia.

Este artigo apresenta um estudo exploratório acerca da proposta de solução. Além desta introdução, o trabalho é composto da seguinte forma: a Seção 2 apresenta o referencial teórico; a Seção 3 descreve a experiência; a Seção 4 apresenta os próximos passos e trabalhos relacionados e a última seção conclui o artigo.

2. REFERENCIAL TEÓRICO

Nesta seção é apresentado o referencial teórico utilizado como base desta pesquisa.

2.1 Group Storytelling

Em seu trabalho, Tobin e Snyman fazem uma revisão da literatura acerca das histórias e do conto de histórias [5]. O artigo apresenta a utilização de histórias em diversas organizações e aponta que todas as organizações, grandes ou pequenas, são dependentes de histórias para seu funcionamento [6].

No contexto dos negócios, os autores definem que as histórias descrevem uma sequência de decisões, ações ou eventos (passados, presentes ou futuros, reais ou imaginários) que envolve personagens (nominados ou não), em uma organização onde o desafio de negócio ou uma oportunidade deve ser endereçada. Elas podem ser contadas em qualquer formato (escrito ou oral) e usar uma variedade de mídias. Já a contagem de histórias (*storytelling*) é um método ou caminho através do qual a história é contada. Pode ser definido como as práticas, ferramentas e os papéis desempenhados pelos envolvidos na comunicação de uma história para uma audiência [5].

Tobin e Snyman expõem a opinião de outros autores que defendem uma série de benefícios do método de contagem de histórias: histórias tendem a ser mais bem fixadas que ideias abstratas; capacita o entendimento de uma maneira relevante; é uma maneira efetiva de capturar o conhecimento; permite a comunicação rápida, natural, clara, confiável, colaborativa, persuasiva, e acurada; permite o compartilhamento de conhecimento numa população diversificada; é carregada de razão e emoção. No trabalho também referenciam o uso prático em empresas americanas (3M e Apple), japonesas (Sony e Toshiba), européias (Club Med e Océ), entre outras [5].

Com relação a histórias contadas de forma coletiva, *group storytelling*, existem outros benefícios. A partir do conceito fundamental de que uma pessoa possui parcela do conhecimento, esta parcela poderá ser complementada com outros participantes que, além da lembrança, poderão contribuir com outras perspectivas. A leitura, comentário e comunicação são capazes de ativar a memória e aumentar a habilidade dos membros de testemunharem a narrativa. Em histórias contadas por equipes de trabalho, cada participante pode interagir e apresentar a sua percepção dos fatos [7].

2.2 Modelagem Conceitual e Ontologias

Modelagem Conceitual pode ser definida quanto à sua função como “em essência, preocupada com o desafio de representar o conhecimento conceitual de uma forma que é adequada e compreensível para todos os *stakeholders* tanto no desenvolvimento quanto no uso destas representações, e independente de qualquer eventual realização técnica de representação” [3]. Descrever um sistema em termos de modelos conceituais significa descrevê-lo constituído de objetos pertencentes a diferentes classes, com propriedades distintas, e relacionados uns com os outros de várias maneiras [1].

Em sistemas de informação, a modelagem conceitual é utilizada no processo de desenvolvimento, proeminentemente nas fases de análise e projeto. Dentro de quatro propósitos maiores da modelagem conceitual apresentados por Recker [3], destacamos o suporte à comunicação entre *stakeholders* e o auxílio aos analistas para o entendimento do mundo real. Modelos conceituais são importantes para a prevenção de erros na fase de requisitos e proveem qualidade nos sistemas de informação a serem entregues [3].

Uma abordagem genérica para modelar conceitualmente um domínio (universo do discurso) é a utilização de gramática de modelagem e métodos de modelagem conceitual para, ao fim, obter um esquema conceitual [8]. Como exemplos de gramáticas, podemos citar a gramática associada ao modelo Entidade Relacionamento (ER) [9] e a Linguagem de Modelagem Unificada (UML) [10]. Já como exemplo de métodos de modelagem, podemos citar a Análise Estruturada [11].

A Engenharia de Ontologia amplifica os objetivos para além da criação e especificação de modelos. Neste caso, partindo de um universo do discurso conhecido, o processo utiliza métodos tais como MethOntology, WebODE ou On-To-Knowledge em conjunção com linguagens ontológicas tais como OWL, ou RDF [8] [12] [13] [14] [15] [16] [17]. A Figura 1 representa os elementos dessas duas abordagens: Gramáticas de Modelagem aliada a Métodos de Modelagem Conceitual obtendo ao fim esquemas conceituais; e a Engenharia de Ontologias cujo resultado são Ontologias de Sistemas de Informação. Não é correto crer que estas abordagens são mutuamente exclusivas; ao contrário, são complementares, pois o uso da Engenharia de Ontologia leva a

modelos conceituais melhores e, diante do que foi dito anteriormente, modelos conceituais melhores levam a sistemas de informação com maior qualidade.

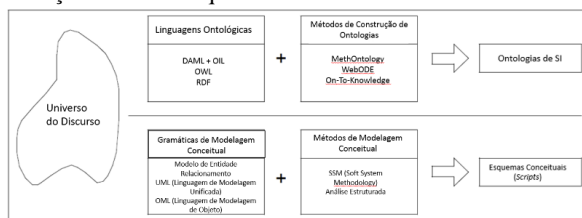


Figura 1. Comparação entre Engenharia de Ontologia e Modelagem Conceitual. (Fonseca, 2007).

Na Filosofia, Ontologia é a descrição básica das coisas no mundo [8]. Guarino refina o significado filosófico de ontologia como sendo um sistema particular de categorias que refletem uma visão particular do mundo [18]. No campo das ciências da computação e de sistemas de informação, a ontologia “refere-se a um artefato de engenharia, constituído de um vocabulário específico usado para descrever uma certa realidade, mais um conjunto de premissas explícitas de acordo com o significado pretendido das palavras do vocabulário” [18]. Mais formalmente:

“Uma ontologia é uma teoria lógica para representar o significado pretendido de um vocabulário formal, ou seja, o seu compromisso ontológico para uma conceptualização particular do mundo. Os modelos pretendidos de uma linguagem lógica usando este vocabulário é restringido pelo compromisso ontológico. Uma ontologia reflete indiretamente este acordo (e a conceptualização subjacente) pela aproximação dos modelos pretendidos.”

(Guarino, 1998)

Diante desta definição, o conceito de ontologia foi adotado no campo dos Sistemas de Informação na busca de respostas encontradas para as dificuldades no processo de modelagem conceitual. “A adoção de ontologias avançou a pesquisa na modelagem conceitual capacitando projetistas de sistemas no desenvolvimento de *scripts* de modelos conceituais melhores” [8]. O uso da ontologia no processo de desenvolvimento habilita os desenvolvedores a um maior nível de reuso do conhecimento, permitindo reuso e compartilhamento do conhecimento de domínio entre sistemas em plataformas heterogêneas [8] [18].

Sistemas de Informação orientados por ontologias (*ontology-driven*) começam com horizontes pessoalmente realizados (isto é, interpretações pessoais com base no conhecimento e na experiência). Esses horizontes são descritos em alguma linguagem, formal ou informal, preocupando-se, primeiro, em obter uma ontologia para o Sistema de Informação. Esta ontologia é, por definição, mais ampla que os esquemas conceituais e podem incluir pontos de vista de uma comunidade maior. São teorias que descrevem e explicam o domínio. Este mesmo domínio é aquele que os Sistemas de Informação tentam modelar. Desta forma, a ontologia provê uma gramática para o modelo conceitual capaz de validar os modelos em relação à realidade [8] [18]. A Figura 2 sintetiza o caminho entre o horizonte pessoal e os fatos na base de dados.

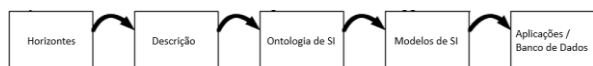


Figura 2. Caminho entre o horizonte pessoal e os fatos na base de dados ou aplicações [8]

2.3 Aprendizado de Ontologias (*Ontology Learning*)

Técnicas de aprendizado de ontologias objetivam apoiar o engenheiro de ontologias na tarefa de criar e manter uma ontologia [19]. Aprendizado de ontologias refere-se à extração de elementos ontológicos (conhecimento conceitual) a partir de uma entrada de informações e a construção de ontologias a partir delas. Ela é composta por um conjunto de técnicas capazes de construir a partir do zero, enriquecer ou adaptar uma ontologia existente de forma automática ou semiautomática utilizando diversos recursos [4] [20] [21].

Aprendizado de ontologias utiliza conhecimentos de diversos campos, tais como: aprendizado de máquina [22], processamento de linguagem natural [23], recuperação de informação [24], inteligência artificial [25] e gerenciamento de banco de dados [4] [21] [26]. Pode ser categorizada de acordo com os tipos de dados de entrada: estruturado, semi-estruturado e não-estruturado.

Dados não-estruturados podem ser textos, tais como livros ou revistas. Dados semi-estruturados podem ser texto em arquivos HTML ou XML. Dados estruturados são dados em bases de dados. Nosso estudo exploratório concentra-se em aprendizado de conceitos a partir de dados não estruturados. Nas referências é possível encontrar mais sobre aprendizado de ontologias em outras categorias de dados [4] [21] [26] [20].

No contexto de aprendizado de ontologias em dados não-estruturados pode haver três categorizações, conforme a abordagem necessária para aprender uma ontologia: abordagem estatística; processamento de linguagem natural; abordagem mista. A primeira abordagem é explicada no trabalho [4] e, em termos gerais, utiliza medidas estatísticas para verificar a representatividade de uma palavra-chave num conjunto de textos e, em seguida, são realizadas novas rodadas de busca nos documentos onde esta palavra foi encontrada para encontrar filhos destes conceitos (palavras próximas ao termo representativo, por exemplo) e, assim, determinar a hierarquia.

A segunda abordagem utiliza um conjunto de padrões sintáticos da linguagem natural para descobrir a relação de dependência entre as palavras. No trabalho de Sabou et al (2005), por exemplo, os passos de extração da ontologia são: análise de dependência, padrões sintáticos, construção de ontologia e poda da ontologia. Para exemplificar, Hazman et al (2011) cita a relação de dependência entre os termos que pode ser reconhecida na frase “encontrar sítios antigênicos em proteína”. O adjetivo “antigênico” modifica o substantivo “sítio” e “sítio é o objeto do verbo “encontrar”. Então, é possível anotar um texto sintaticamente e, utilizando um conjunto de padrões, organizar a informação para construir uma ontologia. Por exemplo, Sabou et al (2005) utilizam essa categoria de padrões. No seu trabalho, um padrão sintático (em inglês) “NN” e “Nmod”, por exemplo, identificava os conceitos. Verbos identificavam funcionalidades. Preposições identificavam relações de parte-todo entre dois termos (*meronymy*, em inglês). Um método automático utilizando esta abordagem é apresentada por Cimiano, Hotho, Staab em 2005 [27].

A terceira abordagem utiliza tanto anotação semântica quanto a análise estatística. Esta abordagem é aplicada no Text2Onto¹, um framework de aprendizado de ontologias a partir de recursos textuais. A Figura 3 apresenta a arquitetura do Text2Onto que é centrada em torno do Modelo Probabilístico de Ontologia

¹ <https://code.google.com/p/text2onto/>

(Probabilistic Ontology Model – POM). O POM armazena o resultado de diferentes algoritmos de aprendizado de ontologia. O POM pode ser traduzido em linguagens de representação de ontologias tais como RDFS², OWL³ e F-Logic⁴. Essa capacidade é devida à uma Biblioteca de Primitivas de Modelagem (MPL, do inglês, Modeling Primitive Library), que define as primitivas das linguagens de forma declarativa e aos escritores de ontologias, que são responsáveis por traduzir as primitivas de modelagem instanciadas em uma das linguagens de representação do conhecimento. Os algoritmos são inicializados por um controlador (Controlador de Algoritmo) que dispara um pré-processamento linguístico dos dados (NLP, do inglês, Natural Language Processing) no Corpus disponível. O controlador executa o algoritmo de aprendizado na ordem adequada no repositório de referência (Base de Dados acessada e controlada pelo gerenciador de referências) e aplica as alterações necessárias no POM. O trabalho de Cimiano e Volker (2005) [28], na Figura 3, explica com maiores detalhes o *framework*.

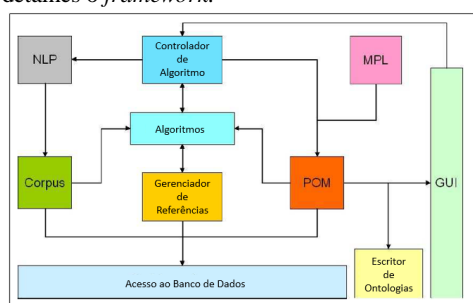


Figura 3. Arquitetura do Text2Onto [28].

2.4 Método de Construção Automática de Ontologias

Existem diversos métodos para construção de ontologias. Podemos citar o OTK [29], Methontology [16] e Diligent [30], por exemplo. Estes três métodos possuem o foco nos engenheiros de ontologias e não em máquinas e, por isso, não são de interesse do presente trabalho, uma vez que o foco de interesse é acerca de abordagens automáticas.

A abordagem automática que foi adotada é baseada na de Bedini e Nguyen (2007) [31], a qual representa um processo de geração automática de ontologia e critérios de avaliação. A Figura 4 apresenta este processo, descrito em detalhes a seguir.

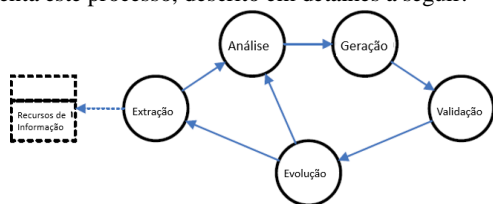


Figura 4. Processo de geração automática de ontologias [31]

Na fase de extração é obtida a informação necessária para geração de elementos da ontologia (conceitos, atributos, relacionamentos e axiomas) a partir de um corpus. Como citado, esta informação pode ser estruturada, semi-estruturada ou não-estruturada. A técnica de extração varia de acordo com essa origem, tais como: processamento de linguagem natural (NLP), clustering,

aprendizado de máquina, semântica, morfológica, léxica ou uma combinação delas.

O passo de análise foca no *matching* da informação recuperada e/ou com a combinação de duas ou mais ontologias existentes, dependendo do caso de uso. Essa etapa requer, além das técnicas da primeira etapa, analisadores semânticos para detectar sinônimos homônimos e outras relações.

A geração é a etapa de *merging* de ontologias e da formalização do meta-modelo usado pela ferramenta que possa ser interpretado por outras aplicações. A formalização se dá com a geração de documentos OWL e RDF-S.

A etapa de validação tem como objetivo retirar conceitos e relacionamentos errados introduzidos nas etapas anteriores. Normalmente feita manualmente, mas em alguns casos pode ser automatizada.

A última etapa, de evolução, é necessária uma vez que a ontologia de um domínio não é estática. Pode haver alteração na quantidade de conceitos, nos relacionamentos e outros parâmetros podem ser acrescentados ou modificados.

3. Descrição do Estudo Exploratório

O estudo exploratório desta pesquisa foi realizado a partir dos textos obtidos de uma ferramenta de *group storytelling*⁵. A ferramenta permite que as histórias sejam contadas de forma assíncrona e distribuída e registradas por escrito. A dinâmica é realizada em três fases. Na primeira, a fase de contar, o grupo participa da contagem da história, incluindo, excluindo e discutindo acerca das informações postas por cada colaborador. Na segunda fase, redigir, a história é redigida por apenas um dos membros e, na última fase, a de conclusão, a história é finalizada e pode ser comentada.

Nesta dinâmica podem coexistir quatro papéis. O do participante, que contribui com a contagem da história. O do facilitador, responsável por observar e estimular o grupo para garantir o sucesso. O do redator, ao qual cabe produzir o texto final, formatado e aperfeiçoado. E, finalmente, o do crítico, que evidencia elementos tácitos do conhecimento explicitado.

As histórias utilizadas nesta pesquisa foram criadas no contexto da disciplina de Gestão de Processos de Negócio, na Universidade Federal onde o estudo foi realizado. Cada grupo de alunos era responsável pelo levantamento de um dos processos existentes na secretaria. Em uma das etapas foi realizada uma sessão de história de grupo onde os participantes contavam as histórias dos processos da levantados. A partir do grupo de histórias cadastrados na ferramenta de *group storytelling* foram explorados os conceitos de aprendizado automático de ontologias.

3.1 Fase 0 - Planejamento

Além do método apresentado no referencial teórico [31] e do *framework* Text2Onto [28], para definir como seria a abordagem deste trabalho foram investigadas a abordagem de outros autores, inclusive com tentativas de reproduzi-las. Por exemplo, a abordagem proposta pela OntoLT [32] foi experimentada mas, nela, a anotação linguística é externa à ferramenta e deveria ser realizada antes da utilização da ferramenta para a extração da ontologia. Chegou-se a anotar um texto semanticamente, utilizando

² <http://www.w3.org/RDF/>

³ <http://www.w3.org/2001/sw/wiki/OWL>

⁴ <http://en.wikipedia.org/wiki/F-logic>

⁵ <http://www.se7ti.com.br:8080/conti/login>

a GATE⁶ e o ferramental disponível pelo Stanford NLP Group⁷, mas ambas anotações eram incompatíveis, e haveria um novo esforço para torná-los compatíveis com a ferramenta. Além disso, a proposta do OntoLT é um plugin para o Protégé que funciona apenas com versões antigas do mesmo.

Também avaliou-se a utilização do OntoGen⁸. Este é um editor semiautomático que combina técnicas de mineração com aprendizado de máquina. É capaz de realizar aprendizado supervisionado e não-supervisionado para sugerir conceitos e nomes de conceitos, bem como a visualização da ontologia e dos conceitos [33]. No entanto, ele apenas apóia a gestão de ontologias. Não é capaz de construí-la de forma automática mas apenas mediante ação humana.

Por fim, utilizou-se a ferramenta Text2Onto. Também nesta foram encontradas dificuldades, ainda em fase de instalação. Existe uma série de incompatibilidades de versões com outros ferramentais, mas ela se mostrou capaz de extrair os construtos de forma totalmente automática. O seu uso será apresentado e avaliado no decorrer deste artigo.

Assim, a abordagem segue, além deste passo de planejamento, outros quatro passos. No segundo passo, foi realizado o pré-processamento da fonte de dados. No terceiro passo foi realizada a extração dos conceitos de forma automática. No quarto, foram avaliados os resultados comparando com a extração manual. No último passo, foram exibidos os conceitos extraídos numa ferramenta de edição de ontologia. A Figura 5 apresenta a abordagem.



Figura 5 – Processo de trabalho proposto

3.2 Fase 1 – Tratamento (Tradução)

As extrações em português e em inglês das histórias estão disponíveis online⁹. Os processos já estavam agrupados em histórias diferentes (vide Figura 6, que exhibe a tela da ferramenta). Cada processo foi, então, traduzido para o inglês e incluído num documento separado. O conjunto dos documentos, em inglês, compôs o *corpus* deste trabalho. A tradução foi realizada utilizando o Google Translate¹⁰.



Figura 6 – Histórias disponíveis na Ferramenta Conti.

3.3 Fase 2 – Extração de Conceitos

Definiu-se, de antemão, os algoritmos a serem utilizados e que o trabalho seria apenas no nível de conceito (concept). O Text2Onto utiliza as anotações realizadas no GATE para os quais substantivos são “Conceitos” e nomes próprios (*proper nouns*) são “Instâncias”.

Como esta pesquisa é exploratória, foram avaliadas apenas as opções disponíveis por padrão na ferramenta. Para extração de conceitos e instâncias utilizou-se o algoritmo TFIDFConceptExtraction. As outras opções disponíveis eram o EntropyConceptExtraction; ExampleConceptExtraction; RTFConceptExtraction. A escolha é justificada pelo fato de que o ExampleConceptExtraction é um algoritmo de exemplo; Tanto o RTFConceptExtraction quanto o EntropyConceptExtraction avaliam a frequência do termo no documento, diferenciando-se apenas pelo método; já o TFIDFConceptExtraction utiliza o RTFConceptExtraction, mas faz a sua avaliação também com relação a um rol de documentos e não apenas em relação a um único documento [28].

Foi executada a extração de conceitos de apenas um documento. Antes da extração, no entanto, foram extraídos os conceitos manualmente para fins de comparação com o resultado final. Aleatoriamente optou-se pelo documento “DeclarationOfCompletionOfCourse”. Na próxima subseção serão apresentados e avaliados o resultado obtido.

3.4 Fase 3 – Avaliação

A Tabela 1 apresenta os conceitos extraídos manualmente e os conceitos extraídos automaticamente. A terceira coluna é a análise se foi realizada uma extração correta (sim) ou (não) quando comparadas as extrações automáticas e manuais.

Tabela 1. Comparação da extração dos conceitos de forma automática e manual.

Conceitos Extraídos Manualmente	Conceitos Extraídos Automaticamente	Análise resultado automático
Application	Application	Sim
application form	application form	Sim
	Completion	Não
Course	Course	Sim
Data	Datum	Sim
Date	Date	Sim
Declaration	Declaration	Sim
Degree	Degree	Sim
	department verify	Não
Department		Não
Direction	Direction	Sim
Document	Document	Sim
	document delivery	Não
End	End	Sim
	Order	Não
Request	Request	Sim
Secretariat	secretariat	Sim
Secretary	Secretary	Sim
	Sign	Não
Stamp	Stamp	Sim
Statement	Statement	Sim
	statement verify	Não
Student	Student	Sim

⁶ GATE – General Architecture Text Engineering (<https://gate.ac.uk/>)

⁷ <http://nlp.stanford.edu/>

⁸ Ontogen.ijs.si

⁹ <https://www.dropbox.com/sh/0gpkqbo5tpggh38/AADGXaRDcfd6jMxAv8M5CdKa?dl=0>

¹⁰ <https://translate.google.com/>

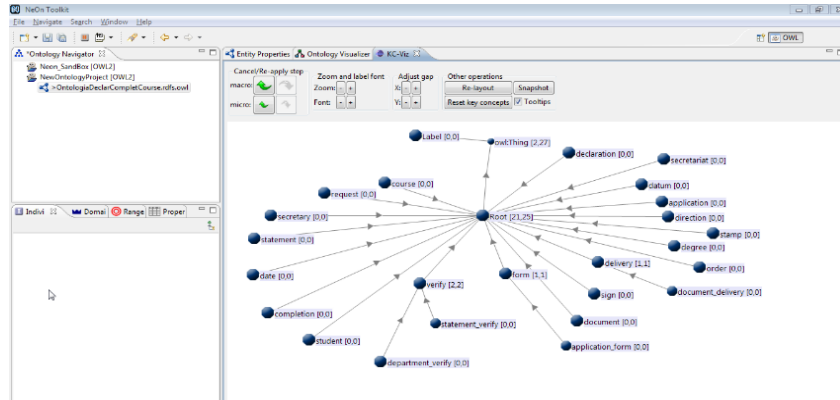


Figura 7 – Conceitos exibidos no NeonToolkit.

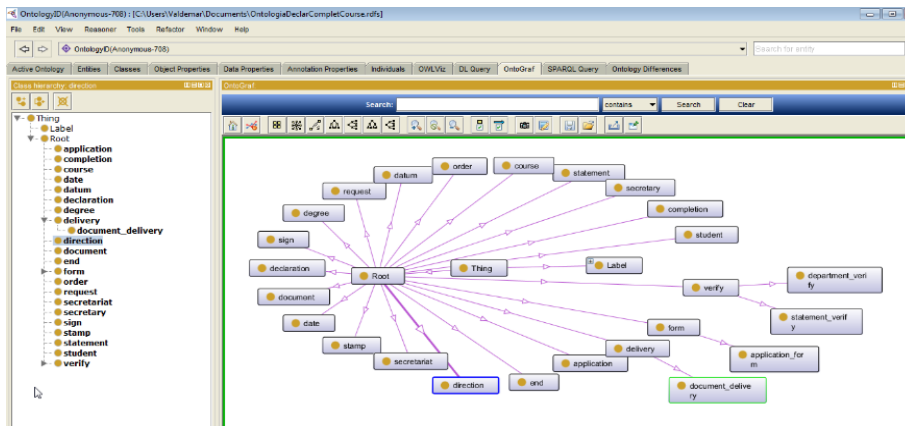


Figura 8 – Conceitos exibidos no Protégé.

3.5 Fase 4 – Exibição

Nesta fase o objetivo é a exibição dos conceitos extraídos utilizando ferramentas de gestão de ontologias. Apesar de a funcionalidade de exportação para OWL não funcionar, realizou-se a exportação para RDF. Em seguida houve a tentativa de exibição em duas ferramentas que são recorrentemente citadas em pesquisas no campo de ontologias: Protégé e NeonToolkit¹¹. A exportação pode ser vista nas Figuras 7 e 8.

Esta fase é importante para observar se há aderência entre o resultado obtido a partir das ferramentas de aprendizado e as ferramentas de construção e manutenção de ontologias. Conclui-se que não houve problemas em relação à exibição. Isto é, aquele que receber a ontologia gerada pelo processo poderá utilizar as ferramentas disponíveis para criticá-la ou evolui-la. Provavelmente isto se deve à maturidade deste ferramental, incluindo questões de sedimentação dos padrões OWL e RDF.

Estas etapas (construção de ontologias e sedimentação de padrões) precedem, tanto em termos de pesquisa quanto em termos práticos, a evolução da etapa de aprendizado de ontologia.

4. PRÓXIMOS PASSOS E TRABALHOS RELACIONADOS

O primeiro passo após a realização deste estudo exploratório é fazer sua expansão. Neste trabalho foi apresentado um passo a passo que possibilitou a extração de conceitos de uma das histórias

disponíveis na ferramenta de *group storytelling*. É possível expandi-lo sob cinco aspectos.

No primeiro aspecto, a expansão é em relação às histórias. Pode-se repetir a mesma abordagem para as demais histórias disponíveis, seja compondo uma ontologia para cada história ou compondo uma ontologia a partir de todas as histórias. Trabalhos de mineração a partir de *group storytelling* já foram realizados no contexto de mineração de processos [35].

No segundo aspecto, a expansão é em relação aos elementos ontológicos. Foram avaliados apenas os conceitos (classes). A abordagem pode ser expandida para a obtenção de instâncias, relacionamentos e categorizações, lidando com outros tipos de desafios. A literatura é vasta neste sentido e já foram citados aqui alguns trabalhos [27] [19] [20].

O terceiro aspecto é quanto à avaliação. A avaliação humana não é o único caminho para verificar os resultados. Técnicas de verificação de correspondência de ontologias pode ser um caminho. Diversos trabalhos buscam avaliar a similaridade com o objetivo de alinhar as ontologias. Como exemplo, é possível citar [36] que é um trabalho completo neste sentido. Nele é apresentado um sistema chamado BLOOMS que é capaz de realizar avaliação de alinhamento de ontologias. Outras abordagens de avaliação de alinhamento são S-Match [37], AROMA [38], RIMoM [39].

O quarto aspecto é em relação a sua comparação com outros idiomas. Existem trabalhos que estendem o aprendizado de

¹¹ http://neon-toolkit.org/wiki/Main_Page

ontologia para outros idiomas, como o chinês [40] ou o português [41], ou mesmo trabalhos *cross-linguísticos* [42] [43].

O quinto aspecto é em relação a ontologias de fundamentação. Leão, Revoredo & Baião [34] buscam obter uma ontologia bem fundamentada utilizando o processamento de linguagem natural aliado aos elementos da Wordnet. Esta abordagem visa obter uma ontologia menos suscetível à ambiguidade de interpretações.

5. CONCLUSÃO

Conclui-se este trabalho pela ciência da importância e da relevância do tema. Este trabalho contribuiu para responder à questão de pesquisa: "Como especificar automaticamente uma ontologia a partir do conhecimento tácito dos envolvidos no domínio?". Contribuir para responder esta questão permitirá auxiliar a gestão e a qualidade da informação.

Foi proposta uma primeira fase onde o conhecimento tácito dos envolvidos no domínio é capturado de forma colaborativa através da técnica de *group storytelling*. Na segunda fase da proposta, foram recuperados os conceitos necessários para confecção da ontologia de forma automática utilizando ferramentas referenciadas em outros trabalhos científicos.

Em 2007, no trabalho de Bedini & Nguyen [31], é feita uma revisão do estado da arte da geração automática de ontologias e os autores concluíram que não havia sistema capaz de gerar automaticamente ontologias. No entanto, este estudo exploratório contribui apontando para a possibilidade em relação a este caminho. Foi verificado, também, que as ferramentas de aprendizado automático de ontologias a partir de textos, talvez pela forte dependência da anotação semântica, ainda não alcançaram o estado da arte que permita a fluência do uso.

Dito isto, trabalhos futuros irão aprofundar a extração de elementos ontológicos. Esta pesquisa indica que, para isso, será necessário estender o estudo, inclusive através de atualização do código das ferramentas disponíveis. Esta percepção é fruto, principalmente, da verificação de descontinuidade das plataformas sobre as quais trabalham as ferramentas exploradas. Por exemplo, o Text2Onto utiliza, só e somente só, a versão 4 do GATE. O GATE, no entanto, está na versão 8.

O presente trabalho evidenciou também a necessidade de, em trabalhos futuros, evoluir em relação à avaliação dos resultados de forma a subtrair a comparação manual. Para isso, é importante um maior aprofundamento no conhecimento em relação a alinhamento ontológico.

Finalmente, outro ponto de melhoria é a comparação de ontologias geradas a partir de descritivos de domínios obtidos de forma individual e de forma colaborativa. As referências de *group storytelling* já apontam as vantagens de obtenção do conhecimento a partir de abordagens colaborativas. No entanto, não foram encontradas referências que façam, especificamente, uma comparação de ontologias extraídas de descritivos de domínio construídos de forma colaborativa de ontologias extraídas de descritivos de domínio de forma individualizada. Assim, concluiu-se que, diante dos resultados obtidos, a tendência de evolução desta pesquisa é no sentido de obter, de forma experimental e não mais em estudo exploratório, os aprimoramentos aqui identificados, além de expandir a abordagem para outros elementos ontológicos.

Encerramos este trabalho destacando, então, o escopo de nossa contribuição: um estudo exploratório que utiliza dados não estruturados obtidos a partir da técnica de *group storytelling* para a extração automática de conceitos de uma ontologia. O objetivo foi avaliar os resultados obtidos a partir da aplicação de um método

para aprender de forma automática uma ontologia a partir das histórias e, assim, contribuir para a gestão e qualidade da informação.

6. AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela FAPERJ (número do projeto: E-26/110.477/2014). Fernanda Araújo Baião e Flávia Maria Santoro são parcialmente financiadas pelo CNPq (número de projeto 309069/2013-0 e 473457/2013-0 respectivamente).

7. REFERÊNCIAS

- [1] P. Johannesson, "The role of business models in enterprise modelling," in *Conceptual modelling in information systems engineering*, Berlin Heidelberg, Springer, 2007, pp. 123-140.
- [2] R. Perret, M. R. Borges and F. M. SANTORO, "Applying group storytelling in knowledge management," in *Groupware: Design, Implementation, and Use*, Springer Berlin Heidelberg, 2004, pp. 34-41.
- [3] J. Recker, *Evaluations of Process Modeling Grammars*, Berlin-Heidelberg: Springer, 2011.
- [4] M. Hazman, S. R. El-Beltagy and A. Rafea, "A Survey of Ontology Learning Approaches," *International Journal of Computer Applications (0975-8887)*, vol. 22, May 2011.
- [5] P. Tobin and R. Snyman, "Once Upon a Time in Africa: a Case Study of StoryTelling for Knowledge Sharing," *Aslib Proceedings*, pp. 130-142, 2008.
- [6] I. Mitroff and R. Kilmann, "Store managers tell: a new tool for organizational problem solving," *Management Review*, vol. 64 No. 7, pp. 18-28, 1975.
- [7] F. M. Santoro and P. Brezillon, "The Role of Shared Context in Group," *Computing and Informatics*, pp. 1001-1026, 2006.
- [8] F. Fonseca, "The double role of ontologies in information science research," *Journal of the American Society For Information Science and Technology*, vol. 58, no. 6, pp. 786-793, 2007.
- [9] P. P.-S. Chen, "The entity-relationship model - toward a unified view of data," in *ACM Transactions on Database Systems (TODS)*, vol. 1, 1976, pp. 9-36.
- [10] J. Rumbaugh, I. Jacobson and G. Booch, *The Unified Modeling Language Reference Manual*, Pearson Higher Education, 2004.
- [11] E. Yourdon, *Modern Structured Analysis*, New Jersey: Englewood Cliffs, 1989, pp. 78-94.
- [12] D. Fensel, "Language Standardization for the Semantic Web: The Long Way from OIL to OWL," in *Distributed Communities on the Web, 4th International Workshop, DCW 2002*, Sydney, Australia, 2002.
- [13] D. Fensel, "Ontologies: a silver bullet for knowledge management and electronic commerce," Berlin, New York: Springer, 2001.
- [14] D. Fensel, F. van Harmelen, I. Horrocks, D. McGuinness and P. Patel-Schneider, "OIL: an ontology infrastructure for the

- Semantic Web," *IEEE Intelligent Systems*, vol. 16, pp. 38-45, 2001.
- [15] A. Gomez-Perez and R. Benjamins, "Knowledge engineering and knowledge management: ontologies and the semantic web," in *13th International Conference, EKAW 2002*, Siguenza Spain, 2002.
- [16] A. Gómez-Pérez, M. Fernández-López and O. Corcho, *Ontology Engineering*, Springer, 2003.
- [17] N. Noy and C. Hafner, "The State of Art in Ontology Design," *AI Magazine*, vol. 18, pp. 53-74, 1997.
- [18] N. Guarino, "Formal ontology in information systems," *Proceedings of the first international conference (FOIS'98)*, vol. 46, 6-8 June 1998.
- [19] S. Staab and R. Studer, *Handbook on Ontologies*, Springer, 2010.
- [20] M. Shamsfard and A. A. Barforoush, "The state of the art in ontology learning: a framework for comparison," *The Knowledge Engineering Review*, vol. 18, no. 4, pp. 293-316, 2003.
- [21] A. Gómez-Pérez and D. Manzano-Macho, "A survey of ontology learning methods and techniques," *Ontoweb Deliverable D1*, 2003.
- [22] T. M. Mitchell, J. G. Carbonell and R. S. Michalski, *Machine Learning: A Guide to Current Research*, Kluwer Academic Publishers, 2011.
- [23] C. D. Manning and H. . Schutze, *Foundations of Statistical Natural Language Processing*, The MIT Press, 1999.
- [24] C. D. Manning, P. Raghavan and H. Schutze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- [25] S. Russel and P. . Norvig, *Artificial Intelligence: A Modern Approach*, 3 edition, Prentice Hall, 2009.
- [26] M. Sabou, C. Wroe, C. Goble and G. Mishne, "Learning Domain Ontologies for Web Services Descriptions: An Experiment in Bioinformatics," in *Proceedings of the 14th international conference on World Wide Web*, Chiba, Japan, 2005.
- [27] P. Cimiano, A. Hotho and S. Staab, "Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis," *JAIR - Journal of AI Research*, vol. 24, pp. 305-339, 2005.
- [28] P. Cimiano and J. Voker, "Text2Onto A Framework for Ontology Learning and Data-driven Change Discovery," *Natural Language Processing and Information*, pp. 227-238, 2005.
- [29] S. Staab, H. Schnurr, R. Studer and Y. Sure, "Knowledge processes and ontologies," *IEEE Intelligent Systems*, 2001.
- [30] C. Tempich, H. Pinto, Y. Sure and S. Staab, "An argumentation ontology for distributed, loosely-controlled and evolving engineering processes of ontologies (DILIGENT)," in *The Semantic Web: Reserach and Applications*, Springer Berlin Heidelberg, 2005, pp. 241-256.
- [31] I. Bedini and B. Nguyen, "Automatic ontology generation: State of the Art," 2007.
- [32] P. Buitelaar, D. Olejnik and M. Sintek, "A Protégé Plug-In for Ontology Extraction from Text Based on Linguistic Analysis," in *Proceedings of the 1st European Semantic Web Symposium (ESWS)*, Heraklion, Greece, 2004.
- [33] B. Fortuna, M. Grobelnik and D. Mladenic, "OntoGen: semi-automatic ontology editor," *Springer Berlin Heidelberg*, pp. 309-318, 2007.
- [34] F. Leão, K. Revoredo and F. Baião, "Learning Well-Founded Ontologies through Word Sense Disambiguation," in *Brazilian Conference on Intelligent Systems*, Fortaleza - CE, 2013.
- [35] J. Gonçalves, F. Santoro and F. A. Baião, "Business Process Mining from Group Stories," in *Proc. 13th International Conference on Computer Supported Cooperative Work in Design*, Santiago, Chile, 2009.
- [36] P. Jain, P. Hitzler, A. Sheth, K. Verma and P. Yeh, "Ontology Alignment for Linked Open," in *Proceedings of the 9th International Semantic Web Conference, ISWC*, Shangai, China, 2010.
- [37] F. Giunchiglia, P. Shvaiko and M. Yatskevich, "S-Match: an algorithm and an implementation of semantic matching," 2004.
- [38] J. David, "AROMA results for OAEI," *Ontology Matching*, p. 122, 2011.
- [39] Z. Wang and et al, "RiMOM results for OAEI 2010," *Ontology Matching*, p. 195, 2010.
- [40] C. Wen, X.-B. Wang and Z.-x. Shi, "Automatic domain-specific term extraction in Chinese domain ontology learning," *Application Research of Computers*, 2009.
- [41] R. Vieira and L. Ribeiro, "Ontolp: Engenharia de ontologias em língua portuguesa," in *Anais do xxviii congresso da SBC-SEMISH-Seminário integrado de software e hardware*, Belem do Pará, 2008.
- [42] P. Vossen, "A multilingual database with lexical semantic networks," Dordrecht, 1998.
- [43] H. Leopold, R.-H. Eid-Sabbagh, J. Mendling, L. G. Azevedo and F. A. Baião, "Detection of naming convention violations in process models for different languages," *Decision Support Systems*, 2013.
- [44] N. Guarino, "Semantic matching: Formal ontological distinctions for information organization, extraction, and integration.," in *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, Berlin, Springer, 1997, pp. 139-170.
- [45] G. Guizzardi, R. Falbo and R. Guizzardi, "A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software," *IEEE América Latina*, vol. 6, pp. 244-251, 2008.