

Anotação de relações semânticas em Objetos de Aprendizagem com o paradigma *Open IE**

Leandro M. P. Sanches¹, Laécio A. Costa¹, Marlo V. S. Souza¹, Laís N. Salvador¹

¹Programa de Pós-graduação em Ciência da Computação (PGCOMP)
Universidade Federal da Bahia (UFBA) – Salvador, BA – Brasil

{leandrompsanches, laeciocosta}@gmail.com, {msouza1, laisns}@ufba.br

Abstract. *The Semantic Web technologies enable the semantic annotation of Learning Objects (LO), which could help the search and retrieval of LO. However, the LO manual semantic annotation is a high-cost, mentally demanding, and, usually, a domain-dependent task. Thus the present work proposes a model for automatic semantic annotation of LO in Portuguese using ontologies and the Open Information Extraction (Open IE) paradigm. We also conducted an empirical study, in which the model implementation was able to annotate semantic relations to LO in Portuguese with a precision of 53.33% and convert them into LOM semantic metadata. The results are promising since the application of Open IE in the semantic annotation of LO is still an unexplored area.*

Resumo. *Os recursos da Web Semântica possibilitam a anotação semântica de Objetos de Aprendizagem (OA) auxiliando a sua busca e recuperação. Como o processo de anotação semântica manual é uma tarefa dispendiosa e geralmente dependente de domínio, o presente trabalho descreve um modelo para anotação automática de relações semânticas em OA utilizando ontologias e o paradigma Open IE. Foi realizado um estudo empírico, no qual a implementação do modelo foi capaz de anotar relações semânticas de OA em português com uma precisão de 53,33% e converter-las em metadados semânticos no padrão LOM. Os resultados são promissores, visto que a aplicação do Open IE na anotação semântica de OA é uma área ainda pouco explorada.*

1. Introdução

Os Objetos de Aprendizagem (OA) são quaisquer recursos digitais que possam ser reutilizados para oferecer suporte à aprendizagem [Wiley 2002]. A quantidade desses recursos tem aumentado nos últimos anos, fato que pode ser atestado pela quantidade crescente de repositórios que armazenam e os disponibilizam [Sanches et al. 2018b].

Uma das principais vantagens dos OA é a possibilidade de reutilização de conteúdo didático digital. Contudo, para ocorrer a reutilização um OA deve estar disponível e acessível para professores e alunos. Neste contexto, segundo Tarouco e Schmitt (2010), os OA apresentam baixa organização no seu armazenamento, o que dificulta a sua pesquisa e recuperação, afetando assim, a disponibilidade destes objetos.

Objetivando solucionar esse problema, na literatura, diversos trabalhos utilizam recursos da Web Semântica. Os recursos da Web Semântica visam estruturar os dados

*Estudo parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

presentes na Web de modo legível por máquina [Berners-Lee et al. 2001]. No contexto educacional, segundo Gonçalves (2007), os recursos da Web Semântica fornecem as ferramentas para a anotação de metadados semânticos em OA, realizado por pessoas e por agentes computacionais, facilitando a sua busca e recuperação.

Os metadados semânticos são um tipo de informação descritiva compreensível por máquina e que possui significado explícito e formal. Os metadados semânticos auxiliam na recuperação de OA facilitando a sua descoberta e reutilização [Wiley 2002, Gonçalves and Carrapatoso 2005].

Segundo Farhat et al. (2015), a anotação manual de metadados semânticos é uma tarefa complexa, demorada e muitas vezes dependente de domínio, necessitando ser refeita caso o OA seja utilizado em outros contextos. Em contrapartida, são poucas as técnicas disponíveis na literatura que realizam a anotação semântica automatizada de OA em língua portuguesa [Sanches et al. 2017].

Neste contexto, o presente trabalho descreve e avalia um modelo independente de domínio para anotação semântica automática de OA em língua portuguesa através de ontologias de domínio. A independência de domínio do modelo de anotação é obtida com a utilização do paradigma *Open Information Extraction (Open IE)* [Wu and Weld 2010] e da parametrização da ontologia de domínio. A utilização do paradigma Open IE para anotação semântica de OA foi anteriormente proposta por [Sanches et al. 2018b], com o foco em OA textuais e sem resultados empíricos. Neste estudo, o modelo de Sanches et al. (2018b) é expandido para ser aplicável a diversos tipos de OA, além do formato textual, e também é realizada uma avaliação empírica.

A maioria dos estudos presentes na literatura realizam somente a anotação semântica de instâncias de entidades, mas não consideram o contexto textual semântico no qual essas entidades estão inseridas [Atkinson et al. 2013]. No modelo apresentado, optou-se pela utilização do paradigma *Open IE*, pois ele possibilita a extração de relações textuais no lugar da identificação de apenas entidades, permitindo assim anotar tanto o significado quanto os relacionamentos entre as entidades.

O restante deste artigo está organizado conforme segue. Na Seção 2 são discutidos os conceitos de anotação semântica e Web Semântica. Na Seção 3 é apresentado o paradigma *Open IE*. A Seção 4 descreve o modelo proposto neste estudo. Na Seção 5 é descrita a sua avaliação. Os trabalhos relacionados são apresentados na Seção 6. Por fim, na última seção estão as considerações finais e os trabalhos futuros.

2. Web Semântica e anotação semântica

A maior parte do conteúdo da Web atual não possibilita que os programas de computador os manipulem de forma significativa. Em contrapartida, a Web Semântica está sendo projetada para ser manipulada tanto por seres humanos quanto por máquinas, pois os seus recursos possibilitam a criação de agentes computacionais capazes de processar o significado atribuído aos dados disponíveis na Web [Berners-Lee et al. 2001].

Entre esses recursos, é possível citar a linguagem *Resource Description Framework (RDF)* e as ontologias. A linguagem RDF tem como objetivo definir um modelo para descrever os conteúdos presentes na Web [Dziekaniak and Kirinus 2004], i.e. um modelo de metadados. As ontologias fornecem os meios para que um agente computaci-

onal determine os significados comuns entre dados de fontes diferentes, possibilitando a comparação e combinação das informações [Berners-Lee et al. 2001].

No contexto educacional, os referidos recursos da Web Semântica possibilitam a anotação semântica de OA, i.e. a adição de metadados com significado explícito, conhecidos como metadados semânticos. Esses metadados não só permitem pesquisas precisas de OA, como também favorecem a combinação de OA [Gonçalves 2007]. Contudo, a anotação manual de metadados semânticos em OA exige demasiado esforço humano e a extração automática de informação de documentos é uma possível alternativa para a geração de anotações semânticas desse tipo de recursos [Nunes and Fileto 2007].

3. Extração Aberta de Informações

A Revisão Sistemática de Literatura de Sanches et al (2017) mostra que para realizar a anotação semântica automatizada de um OA é necessário definir meios de se obter as informações que descrevam esse objeto. Nesse contexto, para a área de Processamento de Linguagem Natural, o conjunto de métodos e técnicas que objetivam extrair dados estruturados a partir de fontes não estruturadas é chamado de Extração de Informação [Etzioni et al. 2008].

Entre os paradigmas de Extração de Informações, um campo de pesquisa emergente é a Extração Aberta de Informações (do inglês, *Open Information Extraction - Open IE*). Os sistemas *Open IE* são independentes de domínio e utilizam padrões genéricos para extrair um grande conjunto de relações semânticas entre conceitos a partir de grandes fontes de dados heterogêneos [Etzioni et al. 2008, Pereira and Pinheiro 2015, Oliveira et al. 2017].

Um extrator *Open IE* pode ser entendido como uma função que dado um documento “d”, extrai um conjunto de tripas (*arg1*, *rel*, *arg2*) para cada relação explicitamente contida no documento, sendo que “arg1” e “arg2” são comumente sintagmas nominais. O parâmetro “rel” é um fragmento textual que indica uma relação semântica entre os argumentos, chamado de descritor de relação. [Wu and Weld 2010].

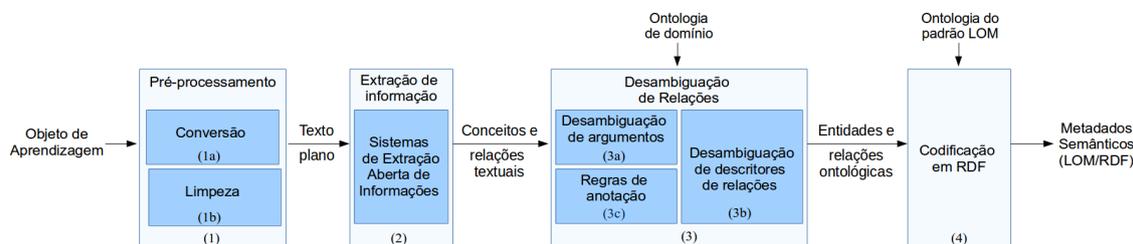
4. Modelo de anotação semântica

Dada a necessidade de automatizar o processo de anotação semântica de OA, Sanches et al. (2018b) propõem um modelo para anotação semântica de OA textuais na forma de relações semânticas. No presente estudo, esse modelo foi expandido para ser genérico o suficiente para a aplicação em diferentes tipos de OA. O modelo de anotação é apresentado na Figura 1, dividido em quatro etapas e prevê a utilização de ontologias de domínio e do padrão *Learning Object Metadata (LOM)* [IEEE Learning Technology Standards Committee - IEEE LTSC 2002].

Um das principais diferenças entre o modelo apresentado por Sanches et al. (2018b) e o modelo aqui descrito encontra-se na primeira etapa. No modelo atual, a primeira etapa (1) Pré-processamento, prevê a conversão do conteúdo do OA em texto, atividade (1a), essa atividade foi elaborada para ser abrangente o suficiente para tornar o modelo de anotação semântica aplicável a diferentes formatos de OA textual, multimídia e hipermídia. Essa conversão pode ser realizada com a aplicação de técnicas como a transcrição de áudio, a conversão de formatos de documentos textuais e o reconhecimento

de caracteres. Subsequentemente, é necessário realizar verificações e limpezas no texto obtido, atividade (1b) da Figura 1.

Figura 1. Modelo proposto de anotação semântica



A etapa (2), Extração de Informações, objetiva identificar os diferentes conceitos e as relações presentes no texto limpo do conteúdo do OA. Devido ao caráter independente de domínio desse modelo, na etapa de Extração de Informações é utilizado um sistema *Open IE*. A saída dessa segunda etapa é um conjunto de triplas (*arg1*, *rel*, *arg2*).

Considerando que a maioria dos sistemas de *Open IE* presentes na literatura não prevê a identificação da semântica das extrações [Sanches et al. 2018a], a etapa (3) do modelo, Desambiguação de Relações¹, utiliza uma ontologia de domínio em língua portuguesa. Essa etapa é responsável por associar os argumentos extraídos do texto com as entidades ontológicas, atividade (3a), e os descritores de relações com as propriedades da ontologia, atividade (3b), fornecendo assim a semântica e possibilitando a legibilidade por máquina dos dados extraídos.

Na terceira etapa do modelo também é prevista a aplicação de diferentes regras utilizadas para a anotação de entidades e relações semânticas, atividade (3c) da Figura 1. As regras de anotação são divididas em dois grupos, dependendo do tipo de informação que é identificada. O primeiro grupo é composto por regras baseadas na desambiguação de descritores de relações, ou seja, que utilizam como entrada uma extração $e_t = (arg_1, rel, agr_2)$ para a qual o descritor de relação *rel* tenha sido desambiguado como instância de uma propriedade ontológica *p*. Assim, uma vez conhecida a propriedade ontológica *p*, o primeiro grupo de regras visa determinar como serão anotados os argumentos da extração e_t .

O segundo grupo de regras é composto por aquelas baseadas na desambiguação dos argumentos e essas regras objetivam obter anotações semânticas em triplas das quais a desambiguação de descritores não foi realizada. Deste modo, dada uma ou mais triplas e_t para as quais foi identificado um conjunto de instâncias ontológicas E_s candidatas a desambiguar os seus argumentos, o segundo grupo de regras é aplicado para determinar propriedades ontológicas *p* que relacionam pares selecionados de entidades $e_i \in E_s$.

Na etapa (4), Codificação em RDF, as relações semânticas anteriormente desambiguadas com a ontologia de domínio são anotadas utilizando o padrão de metadados LOM codificado em RDF através da aplicação do mapeamento do padrão LOM na lin-

¹A Desambiguação de Relação pode ser entendida como a função que associa instâncias relações semânticas *r*, obtidas de uma ontologia, com instâncias de relações (*arg1*, *rel*, *arg2*), extraídas por sistemas OpenIE, que descrevem as mesmas informações que *r* [Sanches et al. 2018a]

guagem RDF de [Rajabi et al. 2014] e da ontologia do padrão de metadados LOM de [Rajabi 2015]. A Figura 2 traz parte de um metadado semântico gerado após a etapa de codificação na avaliação empírica do modelo. Observa-se que esse metadado, através do campo *lom:description*, contém a descrição de uma relação semântica obtida do conteúdo do OA, na qual o celular *Moto G* tem a conexão de dados *HSDPA*. Mais detalhes sobre os OA do estudo empírico e a ontologia utilizada serão apresentados na próxima seção.

Figura 2. Metadado semântico extraído no padrão LOM em RDF

```
<lom:LearningObject
  rdf:about="http://exemplo.org/9ed1f502-ec93-4205-abc8-c70f128b49b6">
  <lom:description>
    <rdf:Description
      rdf:about="http://www.rodriigo.goulart.nom.br/feevale/blogseimagem/Smartphone.owl#moto_g">
      <smart:tem_conexao_de_dados
        rdf:resource="http://www.rodriigo.goulart.nom.br/feevale/blogseimagem/Smartphone.owl#HSPA"/>
      </rdf:Description>
    </lom:description>
  ..
</lom:LearningObject>
```

5. Metodologia da avaliação do modelo de anotação semântica

Objetivando avaliar o modelo de anotação semântica, foi desenvolvido um protótipo em linguagem Java que recebe como entrada OA textuais. As tecnologias para implementação do protótipo foram mantidas as utilizadas por [Sanches et al. 2018b] e para Extração de Informações foi utilizado o extrator *ArgOE* de [Gamallo and Garcia 2015]. Para a avaliação do modelo de anotação, foi realizado um experimento visando determinar a confiabilidade das relações semânticas anotadas pelo protótipo implementado. Para esse experimento, optou-se por utilizar um conjunto de 136 OA na forma de artigos da Wikipédia em português do domínio de celulares inteligentes. Esses artigos, recuperados a partir da categoria *Smartphones*² da Wikipédia, contém em sua maioria informações descrevendo diferentes modelos desses aparelhos. A decisão de utilizar artigos da Wikipédia considerou o grande volume de dados disponível, a licença de uso desses dados e que individualmente estes artigos podem ser considerados como OA, de acordo com a definição de [Wiley 2002].

A execução do experimento foi dividida em duas fases. A primeira fase, chamada de fase de prévia, objetivou discutir e treinar os anotadores humanos sobre o fenômeno analisado. Para isso foi elaborado um guia de anotação semântica e realizada a anotação de um conjunto pequeno de OA, comum a todos os participantes. Após a anotação desses textos, foram levantadas as dificuldades, dúvidas e sugestões dos anotadores. Também foram analisadas as relações anotadas. Essas informações foram utilizadas para atualização do guia de anotação.

A segunda fase, chamada de fase principal, objetivou avaliar a confiabilidade das anotações. Essa fase da avaliação considerou as métricas de precisão (P), *recall* (R) e *F-measure* (F1) calculadas através da comparação manual entre as relações semânticas manualmente anotadas e as anotadas pelos protótipo implementados.

Neste experimento foi utilizada a ontologia de celulares inteligentes [Mertins 2007]. A escolha da ontologia considerou o fato desta possuir um alto número de propriedades ontológicas. Contudo, as diversas modificações ocorridas no

²<https://pt.wikipedia.org/wiki/Categoria:Smartphones>

domínio de celulares inteligentes ao longo dos anos culminaram na necessidade de atualização dessa ontologia objetivando adequá-la a realidade atual do domínio em análise. O processo de atualização da ontologia de celulares inteligentes teve o auxílio de um especialista no domínio.

Além disso, devido ao alto número de diferentes modelos e fabricantes de celulares inteligentes, a criação manual de instâncias dessa ontologia é um processo oneroso. Para tentar superar esse problema, também foi realizado um processo de criação automática de instâncias de classes da ontologia de celulares inteligentes. Essa criação de instâncias ocorreu através da importação de dados da *Dbpedia*³.

6. Resultados e discussões

Em relação à fase de prévia, uma amostra de três OA foi aleatoriamente selecionada e distribuída a sete anotadores humanos. Desses anotadores, somente quatro realizaram o processo de anotação e anotaram 40 relações únicas. Entre essas, foi possível notar a ocorrência de diferentes relações com informações semanticamente similares e também foi possível notar a ocorrência de divergências em como determinar os limites de algumas entidades ontológicas. Devido a essas divergências, na fase principal do experimento foi necessário analisar e agrupar manualmente triplas semanticamente equivalentes.

Note que, como a interpretação de determinada informação é uma tarefa subjetiva, a análise manual das anotações pode representar uma ameaça à validade do experimento. Objetivando superar esse problema e reduzir um possível viés, só foram consideradas como equivalentes as triplas que continham as mesmas informações semânticas.

Na fase principal, foi aleatoriamente selecionada uma amostra de 35 OA. Esses OA foram divididos em conjuntos de cinco e distribuídos a sete anotadores humanos. Somente quatro dos anotadores realizaram o processo de anotação e anotaram 212 relações semânticas. Os resultados obtidos nessa fase da avaliação são mostrados na Tabela 1.

Tabela 1. Resultados da fase principal da avaliação

Regras de anotação	P	R	F1
Grupo 1	0,0%	0,0%	0,0%
Todas as regras	53,33%	3,77%	7,05%

Conforme Tabela 1, o sistema se mostrou pouco adequado na anotação de relações semânticas utilizando somente as regras baseadas na desambiguação dos descritores de relações. Em contrapartida, o sistema apresentou uma precisão de 53,33% quando foram aplicadas todas as regras de anotação. Essas diferenças na precisão podem ser interpretadas como indícios de que a desambiguação de argumentos tem uma considerável importância no processo de anotação.

Diferente da precisão, em todas as combinações de regras, o sistema apresentou baixo *recall* com valor máximo de 3,77%. O baixo *recall* e o fraco desempenho do primeiro grupo de regras podem ser justificados pela presença de relações semânticas aglutinadas em uma única sentença nos textos do experimento. Dificultando, assim, a

³<http://dbpedia.org>

anotação automática. Outra possível razão para o baixo *recall* e o fraco desempenho utilizando somente o primeiro grupo de regras é o erro acumulado proveniente do sistema de extração, incluindo a não extração de informações, a extração de informações incompletas e extração de triplas sintaticamente incorretas.

A não extração de triplas em sentenças com relações semânticas manualmente anotadas pode ser devido a problemas de análise sintática ou a limitações das regras de extração do sistema *Open IE* utilizado [Gamallo and Garcia 2015]. Esse extrator obtém informações através da identificação de relações verbais com a aplicação de heurísticas de extração e da análise sintática da sentença. Assim, caso ocorram problemas na análise sintática, o sistema pode não conseguir extrair relações de uma sentença.

A extração de triplas sintaticamente incorretas resultou na separação de informações necessárias para a desambiguação. Um dos motivos desse problema pode ser a presença de termos em inglês ou de termos específicos do domínio analisado, resultando em erros de análise sintática. A extração de informações incompletas consiste em não extrair fragmentos textuais necessários para anotação da relação semântica e pode ter ocorrido devido a erros de análise sintática.

Como por exemplo da extração de triplas sintaticamente incorretas e da extração incompleta de informações pode-se citar a tripla (*um processador, dual-core, com clock de 1 GHz*) obtida da sentença “Também é o primeiro *smartphone* a apresentar o Nvidia Tegra 2, um processador dual-core com clock de 1 GHz”. Nessa a palavra inglesa *clock* é analisada incorretamente e o sistema também não extraiu a placa de vídeo do modelo de celular inteligente.

Além do erro acumulado, ocorreram problemas de desambiguação. Por fim, para quantificar a interferência de cada um dos problemas acima descritos nos resultados da avaliação, foi manualmente analisada uma amostra contendo as relações semânticas anotadas a 5 Objetos de Aprendizagem. Segundo essa análise, a maior influência no baixo *recall* e no baixo desempenho do primeiro grupo de regras é a não extração de triplas pelo *ArgOE*, seguida pela presença de informações implícitas.

Note que, os resultados deste estudo são em parte devido ao fato que tanto a ferramenta de *Open IE* utilizada [Gamallo and Garcia 2015], quanto o método de desambiguação utilizado [Sanches et al. 2018a] apresentaram desempenho abaixo do esperado. Neste sentido, acredita-se que o uso de outras ferramentas de extração para a língua portuguesa, como o *Linguakit* [Gamallo et al. 2018], ou do uso de métodos do estado da arte em detecção de paráfrases para a língua portuguesa na tarefa de desambiguação, podem melhorar a anotação semântica realizada. Deste modo, como ainda há muito a ser explorado, os resultados aqui apresentados e a área da anotação semântica com o paradigma *Open IE* podem ser considerados como promissores.

7. Trabalhos relacionados

A maioria dos estudos encontrados na literatura que possibilitam anotar de forma automatizada o conteúdo de OA textuais em português não obtém informações diretamente do texto do OA. Assim, em [Behr et al. 2016] a adição de descrições semânticas ocorre por meio da conversão de metadados no padrão Objetos de Aprendizagem Baseados em Agentes (OBAA) codificados em *eXtensible Markup Language* (XML) para metadados

semânticos OBAA em RDF. Já em [Silveira et al. 2013] é apresentada uma arquitetura multiagente na qual a anotação é realizada através de inferências aplicadas sobre ontologias de domínio, de informações preenchidas pelo usuário e de uma ontologia do padrão OBAA. Por fim, em [Romo et al. 2013] também é apresentada uma arquitetura multiagente, na qual a anotação semântica é realizada com a aplicação de múltiplas ontologias de domínio e do cálculo da similaridade semântica entre os metadados anotados pelo usuário e as entidades presentes nas ontologias.

Embora esses três estudos demonstrem que é possível fazer a anotação semântica sem necessariamente obter informações diretamente dos OA, o estudo de [Behr et al. 2016] é limitado a anotar somente OA que possuem metadados OBAA anexados e as arquiteturas [Romo et al. 2013] e [Silveira et al. 2013] são dependentes das informações preenchidas pelo usuário. Em contrapartida, o modelo aqui proposto não necessita da anotação prévia de metadados e a descrição semântica do conteúdo do OA não é condicionada a informações preenchidas pelo usuário.

Diferente dos anteriores, [Santanchè 2007] e [Silva and Santanchè 2009] extraem informações diretamente do OA e propõem metodologias de anotação OA de modo concomitante com a autoria através da análise de marcações pré-definidas aplicadas pelos autores de conteúdo durante a produção e formatação do texto. Nesses dois estudos as marcações utilizadas podem ser consideradas como uma ontologia de aplicação e em [Silva and Santanchè 2009] também é utilizada uma ontologia de domínio que descreve o modelo conceitual *IMS Learning Design* (IMS LD), contudo ambos não realizam a associação do conteúdo extraído com ontologias que descrevam um domínio do conhecimento e a anotação é condicionada ao processo de autoria de um novo objeto.

8. Considerações finais

Neste trabalho, foi abordado o problema da anotação semântica automática de OA. Para isso, foi descrito e avaliado um modelo automático e independente de domínio para anotação semântica de OA. Esse modelo recebe como entrada OA em diferentes formatos e tem como saída metadados semânticos codificados em linguagem RDF. Note que, de acordo com a revisão de literatura realizada no âmbito deste estudo, esse deve ser um dos primeiros modelos a utilizar o paradigma Open IE para anotar OA.

O modelo de anotação utiliza os padrões recomendados para Web Semântica, o padrão de metadados LOM e ontologias de domínio. O diferencial desse modelo é a aplicação da Extração Aberta de Informações, com o intuito de se obter independência de domínio e de extrair relações textuais do conteúdo de diferentes tipos de OA. No modelo de anotação o conceito de Desambiguação de Relações e um conjunto de regras de anotação são aplicados para associar as relações textuais extraídas dos OA com relações de uma ontologia de domínio, possibilitando anotar relações semânticas no lugar de apenas instâncias de entidades ontológicas.

Um protótipo baseado no modelo proposto neste estudo foi implementado e um experimento objetivando determinar a confiabilidade das relações semânticas anotadas foi executado. Neste experimento, o protótipo foi capaz de extrair relações semânticas de OA com uma precisão de 53,33%. Os resultados são promissores, visto que a área de anotação semântica de OA com o paradigma Open IE ainda encontra-se nos seus primeiros passos. Como trabalho futuro, serão investigadas as influências de diferentes ontolo-

gias de domínio e de cada regra de anotação utilizada no processo de anotação semântica. Também serão avaliadas a aplicação do protótipo a outros tipos de OA e a utilização de outros sistemas Open IE.

Referências

- Atkinson, J., Gonzalez, A., Munoz, M., and Astudillo, H. (2013). Web metadata extraction and semantic indexing for learning objects extraction. In *Recent Trends in Applied Artificial Intelligence*, pages 131–140. Springer Berlin Heidelberg, Heidelberg.
- Behr, A., Primo, T., and Viccari, R. (2016). Towards educational metadata interoperability on semantic web. In *27th Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1026–1035, Uberlândia. Sociedade Brasileira de Computação (SBC).
- Berners-Lee, T. J., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific american*, 284(5):29–37.
- Dziekaniak, G. V. and Kirinus, J. B. (2004). Web semântica. *Encontros Bibli: Revista eletrônica de biblioteconomia e ciência da informação*, 9(18):20–39.
- Etzioni, O., Banko, M., Soderland, S., and Weld, D. S. (2008). Open information extraction from the web. *Communications of the ACM*, 51(12):68–74.
- Farhat, R., Jebali, B., and Jemni, M. (2015). Ontology based semantic metadata extraction system for learning objects. In *Emerging Issues in Smart Learning*, pages 247–250. Springer Berlin Heidelberg, Heidelberg.
- Gamallo, P. and Garcia, M. (2015). *Multilingual Open Information Extraction*, pages 711–722. Springer International Publishing, Cham.
- Gamallo, P., Garcia, M., Piñeiro, C., Martinez-Castaño, R., and Pichel, J. C. (2018). Linguakit: A big data-based multilingual tool for linguistic analysis and information extraction. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, volume 15, pages 239–244.
- Gonçalves, V. M. B. (2007). *A Web Semântica no contexto educativo: um sistema para a recuperação de objectos de aprendizagem baseado nas tecnologias para a Web Semântica, para o e-learning e para os agentes*. PhD thesis, University of Porto, Porto.
- Gonçalves, V. M. B. and Carrapatoso, E. M. (2005). Um sistema de e-learning para a web semântica baseado na tecnologia de agentes. In: *Conferência IADIS Ibero-Americana WWW/Internet*.
- IEEE Learning Technology Standards Committee - IEEE LTSC (2002). IEEE standard for learning object metadata. *IEEE Std. 1484.12.1-2002*, pages 1–40.
- Mertins, F. (2007). Ontologia de domínio na análise de blogs. Monografia, Universidade Feevale, Novo Hamburgo.
- Nunes, A. M. and Fileto, R. (2007). Uma arquitetura para recuperação de informação baseada em semântica e sua aplicação no apoio a jurisprudência. *Escola Regional de Banco de Dados (ERBD)*, 3.

- Oliveira, L. S., Glauber, R., and Claro, D. B. (2017). Dependente: An open information extraction system on portuguese by a dependence analysis. In *Encontro Nacional de Inteligência Artificial e Inteligência Computacional*, Uberlândia. SBC.
- Pereira, V. and Pinheiro, V. (2015). Report-um sistema de extração de informações aberta para língua portuguesa. In *10th Brazilian Symposium in Information and Human Language Technology*, pages 191–200, Natal. SBC.
- Rajabi, E. (2015). Lom ontology: a vocabulary for exposing ieee lom metadata as linked open data. <http://lov.okfn.org/dataset/lov/vocabs/lom>.
- Rajabi, E., Sicilia, M.-A., Ebner, H., Palmer, M., and Sanchez, S. (2014). Recommendation on exposing ieee lom as linked data 1.0 (second version). http://data.organic-edunet.eu/ODS_LOM2LD/ODS_SecondDraft.html.
- Romo, B. S., Guilherme, I. R., and Queiroz, J. (2013). Uma arquitetura multiagente para sistemas web semântico para gestão de conteúdos educacionais. In *2th Congresso Brasileiro de Informática da Educação (CBIE)*, pages 561–570, Campinas. SBC.
- Sanches, L. M. P., Cardel, V. S., Machado, L. S., Souza, M. V. S., and Salvador, L. N. (2018a). Disambiguating Open IE: identifying semantic similarity in relation extraction by word embeddings. In *Computational Processing of the Portuguese Language*. Springer International Publishing.
- Sanches, L. M. P., Costa, L. A., Salvador, L. N., and Souza, M. V. S. (2018b). Anotação semântica automática de objetos de aprendizagem textuais em português com o paradigma open ie. In *29th SBIE*, volume 29, Fortaleza. SBC.
- Sanches, L. M. P., Costa, L. A., Souza, M. V. S., and Salvador, L. N. (2017). Anotação semântica automática de objetos de aprendizagem: Um mapeamento sistemático de literatura. In *28th SBIE*, Recife. SBC.
- Santanchè, A. (2007). Otimizando a anotação de objetos de aprendizagem através da semântica in loco. In *18th SBIE*, pages 452–461, São Paulo. SBC.
- Silva, L. A. M. and Santanchè, A. (2009). Arara: Autoria de objetos digitais complexos baseada em documentos. In *20th SBIE*, Florianópolis. SBC.
- Silveira, E. L., Galão, M. C., and Gluz, J. C. (2013). Uma ferramenta para fornecer apoio a catalogação de metadados de objetos de aprendizagem - linnaeus. In *24th SBIE*, pages 687–696, Campinas. SBC.
- Tarouco, L. M. R. and Schmitt, M. A. R. (2010). Adaptação de metadados para repositórios de objetos de aprendizagem. *Revista RENOTE*, 8(2).
- Wiley, D. A. (2002). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In *The Instructional Use of Learning Objects*. Agency for Instructional Technology and Association for Educational Communications and Technology.
- Wu, F. and Weld, D. S. (2010). Open information extraction using wikipedia. In *48th Annual Meeting of the Association for Computational Linguistics*, pages 118–127, Stroudsburg. Association for Computational Linguistics.