

Alinhamento entre Ontologias de Topo e de Domínio usando WordNet

Rafael Basso*, Daniela Schmidt*, Cassia Trojahn†, Renata Vieira*

¹*Pontifícia Universidade Católica do Rio Grande do Sul (Brasil)
rafael.basso, daniela.schmidt@acad.pucrs.br, renata.vieira@pucrs.br

† Institut de Recherche en Informatique de Toulouse (France)
{cassia.trojahn}@irit.fr

Abstract. *Matching domain and top-level ontologies is an important task but still an open problem in the ontology matching field, particularly due to their different levels of abstraction. In this paper we propose an approach for matching domain and top ontologies that exploits the background knowledge from WordNet as an intermediate layer. We evaluate our approach using two well-known top-level ontologies, namely DOLCE and SUMO, and one domain ontology from the OAEI conference data set.*

Resumo. *Alinhar ontologias de domínio e topo é uma tarefa importante, mas ainda é um problema em aberto na área de alinhamento de ontologias, principalmente devido aos diferentes níveis de abstração que estas ontologias representam. Este artigo aborda o alinhamento entre ontologias de topo e de domínio, explorando o conhecimento do WordNet como uma ponte semântica entre ontologias de diferentes níveis. Para avaliar a nossa abordagem, nós consideramos duas ontologias de topo bem conhecidas, DOLCE e SUMO, e uma ontologia de domínio do conjunto de dados de avaliação do OAEI.*

1. Introdução

Ontologias podem ser classificadas de acordo com seu nível de generalidade. As ontologias de topo descrevem conceitos genéricos, como, por exemplo, espaço e tempo, ou seja, conceitos básicos, universais e comuns a qualquer domínio. Dessa forma, garantem expressividade e generalidade para uma vasta área de domínios. Por outro lado, as ontologias de domínio descrevem um campo específico relacionado a um domínio, como por exemplo, a aeronáutica ou a biologia. Ontologias de topo executam um papel importante no desenvolvimento de ontologias sendo utilizadas como um guia para a modelagem de ontologias de domínio, ou para incrementar a expressividade e o escopo de ontologias de domínio existentes. Nesse contexto, alinhar ontologias de domínio e topo surge como uma área de pesquisa com potencial de expansão, visto que essa tarefa tem sido executada em sua maioria, de forma manual, conforme as abordagens propostas na literatura.

De acordo com [Euzenat and Shvaiko 2007], o processo de encontrar correspondências entre entidades de duas ontologias é conhecido como *ontology matching*. Esse processo obtém como saída um conjunto de correspondências que é chamado de alinhamento. O alinhamento entre ontologias de topo e de domínio enfrentam desafios, em particular devido ao seus diferentes níveis de abstração. Assim, este artigo apresenta uma

abordagem de alinhamento automático entre ontologias de topo e de domínio, baseando-se no WordNet e na noção de contexto, que tem sido explorado de diferentes maneiras no processo de correspondência ontológica [Maedche and Staab 2002, Wang 2011, Schadd and Roos 2012, Djeddi and Khadir 2014, Schmidt et al. 2016b]. Para avaliar o nosso protótipo, foram executados experimentos utilizando uma ontologia de domínio de conferências e duas ontologias de topo (DOLCE e SUMO). O conjunto de correspondências resultante foi avaliado manualmente.

O restante do artigo é organizado como segue. §2 introduz ontologias de topo, o processo de alinhamento entre ontologias e a estrutura do Wordnet. §3 descreve a abordagem proposta. §4 apresenta os experimentos e resultados alcançados. §5 discute trabalhos relacionados e, por fim, §6 conclui o artigo e apresenta direções para trabalhos futuros.

2. Fundamentação Teórica

Nessa seção introduzimos os principais tópicos que serão utilizados como base teórica para o desenvolvimento do nosso trabalho.

2.1. Ontologias de Topo

Ontologias de topo descrevem conceitos genéricos e universais, garantindo generalidade e expressividade para uma grande área de domínios. Na literatura, diversas ontologias de topo têm sido desenvolvidas, tais como a SUMO (*Suggested Upper Merged Ontology*) [Niles and Pease 2001] e a DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) [Gangemi et al. 2002] (utilizadas neste trabalho). Porém muitas outras foram desenvolvidas (detalhes em, [Mascardi et al. 2007], os quais fazem a análise de diferentes ontologias de topo). Abaixo, introduzimos as ontologias de topo consideradas neste trabalho.

A DOLCE [Gangemi et al. 2002] foi desenvolvida por Nicola Guarino e sua equipe no LOA (*Laboratory for Applied Ontology*). O foco da DOLCE é compreender as categorias subjacentes das tarefas cognitivas humanas e o ambiente sociocultural. Inclui conceitos como *abstract quality*, *abstract region*, *physical object*, *process*, etc. É implementada em *First Order Logic*, KIF (*Knowledge Interchange Format*) e OWL (*Web Ontology Language*).

A SUMO [Niles and Pease 2001] foi desenvolvida pelo *Standard Upper Ontology Working Group*, um grupo de colaboradores do IEEE dos campos de engenharia, filosofia e ciência da informação. É utilizada para pesquisas e aplicações em áreas como linguística e raciocínio. Inclui conceitos como *abstract*, *human*, *social role*, etc. É implementada em SUO-KIF (*Standard Upper Ontology Knowledge Interchange Format*) e existe uma tradução em OWL (*Web Ontology Language*).

2.2. Alinhamento entre Ontologias

De acordo com [Euzenat and Shvaiko 2007], um alinhamento é definido como um conjunto de correspondências entre duas ontologias. O alinhamento é o resultado de um processo de *matching* de ontologias. A definição formal de correspondência dada por [Euzenat and Shvaiko 2007] é: Uma tupla $\langle e_s, e_t, r, n \rangle$, tal que: e_s e e_t são entidades (e.g., conceitos, propriedades, instâncias) de o_s e o_t (*ontologias fonte e alvo*), respectivamente; r é uma relação entre duas entidades e_s e e_t , (por exemplo, *equivalence*, *subsumption*,

disjointness, overlapping); e n é um número de medida de confiança no intervalo [0;1]. A medida de confiança atribui um grau de confiabilidade para a correspondência encontrada no processo de matching.

2.3. WordNet e seu Mapeamento para Ontologias de Topo

WordNet [Miller 1995] é uma base lexical do inglês composta por substantivos, verbos, adjetivos e advérbios, que são agrupados em conjuntos de sinônimos cognitivos, chamados de *synsets*. Em outras palavras, um *synset* é um grupo de termos que tem um significado comum dentro de um certo contexto. O WordNet também fornece descrições textuais dos *synsets*, chamadas de *gloss* que contêm exemplos e definições dos mesmos. Como exemplo, temos o conceito *Poster*, representado pelo *synsets* que agrupa os sinônimos *poster, posting, placard, notice, bill, card*, com a descrição “*a sign posted in a public place as an advertisement; a poster advertised the coming attractions*”. Os *synsets* estão interligados por meio de relações conceituais-semânticas e lexicais. Assim, além da sinonímia, o WordNet inclui relações de super e subordinação, ou hiponímia e hiperonímia.

Devido a essas características, o WordNet é semelhante a uma ontologia [Gangemi et al. 2002], onde relações de hiponímia são como relações de subsunção para os conceitos. Entretanto, o WordNet, para ser considerado como uma ontologia, deveria poder ser interpretado de acordo com uma semântica formal que nos diz algo sobre o mundo e não apenas sobre a linguagem. Para isso [Gangemi et al. 2002] apresentam esforços para alinhar o WordNet com a DOLCE. Da mesma forma, também é possível encontrar alinhamentos com outras ontologias de topo, como a SUMO [Niles and Pease 2003].

O alinhamento entre o WordNet e a DOLCE apresentado por [Gangemi et al. 2002] considera a taxonomia dos substantivos da versão 1.6 do WordNet. Os autores assumem que a relação de hiponímia poderia ser mapeada com a relação de subsunção e a noção de *synset* poderia ser mapeada com a noção de conceito. Em seguida, os conceitos nomeados foram normalizados para obter um nome distinto para cada *synset*. Assim, se um termo tiver um único significado, então seu *synset* é unificado e usado como nome do conceito (*e.g world_wide_web__web__www*). Se o termo for polissêmico, então os diferentes conceitos serão enumerados (*e.g window_1, window_2*) para os *synsets* com a descrição “*a rectangular part of a computer screen that contains a display different from the rest of the screen*” e “*a framework of wood or metal that contains a glass windowpane and is built into a wall or roof to admit light or air*”, por exemplo.

Na mesma linha, [Niles and Pease 2003], desenvolveram um alinhamento entre a ontologia de topo SUMO e o WordNet. Os autores também utilizaram a versão 1.6 do WordNet. A metodologia utilizada consiste em identificar relações de sinonímia, hiperonímia e instanciação de conceitos da SUMO com *synsets* do WordNet. Um exemplo da estrutura do alinhamento do WordNet com a SUMO, indicando uma relação de equivalência (=) pode ser conferido a seguir:

```
00008864 03 n 03 plant 0 flora 0 plant_life 0 027 @ . . . — a living organism lacking the  
power of locomotion &%Plant=
```

Os primeiros 8 dígitos indicam o ID do *synset*, as três palavras seguintes são as palavras que compõem o *synset*, a definição do *synset* segue logo após a barra horizontal e por último o conceito da SUMO e o tipo de relação sucedem o prefixo &% .

3. Abordagem Proposta

Neste trabalho, apresentamos o desenvolvimento de uma abordagem para alinhamento de ontologias de topo e de domínio. Nossa abordagem utiliza os alinhamentos existentes entre as ontologias de topo DOLCE e SUMO e WordNet. Dessa forma, o WordNet atua como uma camada intermediária entre a ontologia de topo e a ontologia de domínio. Um protótipo foi implementado utilizando a linguagem Java, e recursos como *Ontology API*, *Alignment API* e WordNet.

O processo de matching divide-se em duas fases. A primeira fase, de identificação do *synset*, é composta por três etapas: a constituição do contexto; o processamento do contexto; e a seleção do *synset* adequado. Essas etapas ocorrem para cada conceito da ontologia de domínio. A segunda fase é responsável por utilizar o *synset* recuperado na primeira fase e identificar o mapeamento do mesmo com a ontologia de topo. Esta fase é diferente para as ontologias de topo DOLCE e SUMO. Conforme descrito na seção 2.3, os alinhamentos entre o WordNet e as ontologias de topo são modelados de maneira diferente. Para DOLCE, os autores desenvolveram uma ontologia em formato OWL composta pelos conceitos na ontologia de topo e o conceito que representa o *synset* do WordNet devidamente mapeados. Já para a SUMO, os autores desenvolveram um arquivo texto que descreve o *synset* com seu respectivo alinhamento com o conceito de topo. Assim, implementamos o nosso protótipo considerando os diferentes formatos de alinhamento entre as ontologias e o WordNet. Os passos do mapeamento serão detalhados a seguir:

3.1. Fase 1 : Constituição do Contexto

Essa etapa foi sub-dividida em três tarefas: a extração dos conceitos; a extração da anotação do conceito; e a extração dos conceitos de níveis superiores e inferiores. Após a realização dessas três etapas, o processo cria um contexto pré-processado, ou seja, composto por todos os elementos recuperados incluindo termos repetidos, compostos e sentenças. O processo inicia carregando a ontologia de domínio. Logo após, extrai-se os conceitos da ontologia e, para cada conceito extraído recupera-se sua descrição, o seu nome e o seus vizinhos (super-conceitos e sub-conceitos). Adicionando essas informações no contexto de cada conceito. Todo conceito extraído é armazenado em uma lista, junto com seu contexto.

Um exemplo de um conceito extraído, indicando a estrutura do seu contexto é apresentado a seguir:

- Conceito: *Committee_member*;
- Super-conceitos: *Person* (Neste caso há apenas um super-conceito, entretanto pode variar a quantidade);
- Sub-conceitos: *Chair* e *Co-chair*;
- Anotação (descrição): Nulo (Neste caso a ontologia não fornece descrições para seus conceitos, sendo assim, a descrição é nula).

Com base nessas informações, a estrutura do contexto fica da seguinte forma:

- Contexto: *Committee_member, Person, Chair, Co-chair*

O processo de constituição de contexto é implementado conforme descrito no Algoritmo 1. O contexto é construído apenas para os conceitos das ontologias (as propriedades não são consideradas).

```

Input: ontologia ( ontologia de domínio );
for ( OWLClass cls IN getConcepts(ontologia) ) do
    Conceito cnp = new Conceito( );
    List<String> context = new List( );
    if ( ! isAnonymous( cls ) ) then
        cnp.setConceito( cls );
        context.add( cls.getID( ) );
        context.add( cls.getAnnotation( ) );
        extractSuperConceitos( cls, context );
        extractSubConceitos( cls, context );
        cnp.setContext( context );
        listaConceitos.add( cnp );
    end
end

```

Algoritmo 1: Constituição do contexto

3.2. Fase 1 : Processamento do Contexto

Nessa etapa, a entrada é o contexto de cada conceito. Primeiro, removemos os caracteres especiais, em seguida todos os termos compostos e sentenças são transformados em termos simples (função *singleTerms*, modifica a lista passada por parâmetro). A seguir todos os termos repetidos são removidos (função *rmRepeatedW*, modifica a lista passada por parâmetro) e por último removemos as *stopwords*.

Um exemplo do processamento do contexto, dando continuidade ao exemplo anterior pode ser visualizado a seguir:

- Contexto pré-processado: *Committee_member, Person, Chair, Co-chair*
- Primeiro passo: *Committee member, Person, Chair, Co chair*
- Segundo passo: *Committee, member, Person, Chair, Co*

O processo de processamento do contexto pode ser visualizado no Algoritmo 2.

```

for ( Conceito cnp IN listaConceitos ) do
    List < String > cntxt = new List( );
    cntxt = cnp.getContext( );
    List<String> temp = new List( );
    if ( cntxt != NULL ) then
        temp = rmEspecialChar( cntxt );
        singleTerms( temp );
        rmRepeatedW( temp );
        cntxt.clear( );
        cntxt = temp.rmStopW( );
        cnp.setContext( cntxt );
    end
end

```

Algoritmo 2: Processamento do contexto

3.3. Fase 1 : Seleção do Synset Adequado

Conceitos representados em ontologias podem apresentar diferentes significados, ou seja, ser polissêmicos, variando de acordo com o domínio que os mesmos estão representando. Para identificar o significado que o conceito visa representar, nos baseamos no contexto construído para o mesmo e no uso do WordNet como um recurso externo auxiliar na identificação dos diferentes significados que um conceito pode representar. Assim, identificado um conjunto de *synsets* correspondentes no WordNet, utiliza-se o contexto para selecionar o *synset* que melhor expressa o conceito. A seguir, essas etapas serão detalhadas.

Inicialmente ocorre o acesso ao WordNet e passa-se por parâmetro o nome do conceito quando trata-se de conceito simples. Para conceitos com nomes compostos, adotou-se apenas a última palavra do conceito no acesso ao WordNet (e.g. *Committee member*, busca-se apenas *member* no WN). Caso o nome do conceito composto tenha um adjetivo, esse adjetivo é removido, e assim utiliza-se a palavra restante.

Então é recuperado o *IDictionary* (função *recoverDictionary*). Com o *dictionary* aberto recupera-se um *Index* passando por parâmetro o nome do conceito juntamente com um atributo que indica que a palavra passada por parâmetro é um substantivo.

```
IDictionary dict = recoverDictionary();
dict.open( );
for ( Conceito cnp IN listaConceitos ) do
    List<String> context = cnp.getContext( );
    IIndexWord idxWord = dict.getIndexWord( cnp.getName( ),
        POS.NOUN);
    if ( idxWord != NULL ) then
        int max = 0;
        List<ISynset> listaSynset = idxWord.getISynsets( );
        for ( ISynset synset IN listSynset ) do
            List<String> synsetContext = recoverSynsetContext( wordId );
            int size = intersection( context, synsetContext );
            if ( size > max ) then
                max = size; cnp.setGoodSynset( synset );
            end
        end
    end
end
dict.close( );
```

Algoritmo 3: Seleção do Synset adequado

Se esse *Index* for diferente de nulo, então ocorrerá a recuperação do *synset*. Entretanto, um conjunto de *synsets* (*listSynsets*) será recuperado e para selecionar o correto é necessário criar um contexto para o *synset*, formado pelas palavras que o compõem e pelo seu *gloss* (definição do *synset*). Após criado esse novo contexto, compara-se ele com o contexto do conceito (função *intersection*), o contexto dos *synsets* com o maior número de elementos iguais aos elementos do contexto do conceito tem seu *synset* selecionado como

adequado para o conceito. O processo de seleção do *synset* adequado pode ser visualizado no Algoritmo 3. Como mencionado anteriormente, a segunda etapa da abordagem é diferente para as ontologias DOLCE e SUMO, que será descrito nas próximas seções.

3.4. Fase 2 : Alinhamento entre Ontologias de Domínio e SUMO

Para o alinhamento entre a ontologia de domínio com a ontologia SUMO, nossa abordagem utiliza o alinhamento prévio entre o WordNet e a SUMO descrito em 2.3.

Para cada conceito já extraído da ontologia de domínio, foram utilizados o ID do *synset* já selecionado como adequado. Após selecionar o ID, percorre-se o arquivo *WordNetMapping-nouns* (função *findAli*), para encontrar o conceito alinhado com a SUMO. Por último utiliza-se a classe *Mapping* para gerar um arquivo de saída em formato .rdf, contendo como atributos *source* (nome do conceito de domínio), *target* (nome do conceito de topo), *relation* (tipo de relação) e o *measure* (informa a confiabilidade dos dois conceitos). Utiliza-se a mesma função (*findAli*) para selecionar o *target*, *relation* e *measure*. Passa-se o *Mapping* de cada conceito para um função que gera o arquivo de saída em RDF. O processo de alinhamento com a ontologia de topo SUMO pode ser visualizado no Algoritmo 4.

```
for ( Concept concept IN listConcept ) do
    Mapping map = new Mapping( );
    String code = concept.getGoodSynset( ).getID( );
    map.setSource( concept.getName( ) );
    findAli( code, map );
    preRDF( map );
end
outRdfFile();
```

Algoritmo 4: Alinhamento com a ontologia de topo SUMO

3.5. Fase 2 : Alinhamento entre Ontologias de Domínio e DOLCE

Essa fase utiliza o alinhamento prévio entre a ontologia de topo DOLCE e o WordNet. Para cada conceito previamente extraído da ontologia de domínio, identifica-se (a partir do *synset* anteriormente desambiguado) o conceito correspondente na ontologia que representa o alinhamento entre o WordNet e a DOLCE. Essa identificação ocorre de duas maneiras: a primeira ocorre pelos termos que compõem o *synset* (função *compareW*), que é comparado com os termos que compõem o nome do conceito do WordNet no mapeamento com a DOLCE. E a segunda ocorre pelo *gloss* do *synset* (função *compareG*), comparando-o com a anotação do conceito do WordNet no mapeamento com a DOLCE. Após encontrar o conceito equivalente ao *synset*, extrai-se o conceito de nível mais alto a partir dele (função *extractSuper*), esse conceito de nível mais alto equivale ao conceito da DOLCE. A instância da classe *Mapping* é utilizada para gerar a saída. A saída dessa fase é o próprio alinhamento em formato RDF. O processo de alinhamento com a ontologia de topo DOLCE pode ser visualizado no Algoritmo 5.

4. Experimentos

Com o propósito de avaliar nossa abordagem, executamos experimentos envolvendo uma ontologia do domínio de conferências (ontologia *conference*), que faz parte do jogo de

```

Input: onto2 ( wnNounsyn-v7.owl );
for ( Conceito cnp IN listaConceito ) do
    Mapping map = new Mapping( );
    Conceito cnp2 = new Concept( );
    Conceito cnp3 = new Concept( );
    ISynset synset = cnp.getGoodSynset( );
    List<String> synsetWords = getWords( synset );
    map.setSource( cnp.getName( ) );
    cnp2 = compareW( synsetWords, onto2 );
    if ( cnp2 == NULL ) then
        String gloss = synset.getGloss( );
        cnp2 = compareG( gloss, onto2 );
    end
    cnp3 = extractSuper( cnp 2 );
    map.setTarget( cnp3.getName( ) );
    map.setRealtion( > );
    map.setMeasure( true );
    preRDF( map );
end
outRdfFile();

```

Algoritmo 5: Alinhamento com a ontologia de topo DOLCE

Tabela 1. Resultado da avaliação manual das correspondências entre a ontologia de conferência e as ontologias de topo DOLCE e SUMO

Avaliação Manual	Conferência e DOLCE		Conferência e SUMO	
Corretos	26	43,33%	35	58,34%
Incorretos	27	45,00%	17	28,33%
Nulos	7	11,67%	8	13,33%

testes *Conference* das campanhas de avaliação OAEI¹. A ontologia é composta por 60 conceitos. No primeiro experimento, avaliamos o nosso algoritmo na tarefa de alinhar a ontologia de domínio com a ontologia DOLCE. Nesse experimento, dos 60 conceitos de domínio, para 7 não foram encontrados os *synsets* no WordNet retornando como nulo (11,67%). Em 26 casos o alinhamento resultante foi considerado correto na avaliação manual (43,33%) e para 27 casos, o alinhamento não foi considerado correto (45%).

Posteriormente, executamos o mesmo experimento, porém na tarefa de alinhar a ontologia de domínio com a ontologia de topo SUMO. Nesse experimento, dos 60 conceitos de domínio, para 8 não foram encontrados os *synsets* no WordNet retornando como nulo (13,33%). Em 35 casos o alinhamento resultante foi considerado correto na avaliação manual (58,34%) e para 17 casos, o alinhamento não foi considerado correto (28,33%). A Tabela 1 sumariza os resultados da avaliação manual envolvendo as correspondências identificadas pelo protótipo entre a ontologia de domínio de conferências e as ontologias de topo DOLCE e SUMO.

¹<http://oaei.ontologymatching.org/>

As Tabelas 2 e 3 apresentam um fragmento das correspondências encontradas entre os conceitos da ontologia de domínio (coluna 1) e as ontologias de topo DOLCE (coluna 2) e SUMO (coluna 2) juntamente com o resultado da avaliação manual (coluna 3). Os conceitos estão representados em negrito juntamente com sua hierarquia na ontologia. Na avaliação manual, *Y* indica que a correspondência foi considerada correta pelos avaliadores, *N* indica que a correspondência não foi considerada correta e *null* indica que não foram recuperados *synsets* no WordNet para o conceito de domínio, logo, não foi possível determinar o alinhamento com o conceito de topo via WordNet.

Tabela 2. Fragmento das correspondências geradas entre a ontologia de domínio e a ontologia de topo DOLCE

Conceitos de Conference	DOLCE	Avaliação Manual
Person/Conference_participant/ Active_conference_participant	particular/spatio-temporal-particular/ endurant/non-physical-endurant/ non-physical-object/social-object/ agentive-social-object/agentive-figure/ socially-constructed-person	Y
Person/Conference_contributor/ Invited_speaker	particular/spatio-temporal-particular/ endurant/non-physical-endurant/ non-physical-object/social-object/ agentive-social-object/agentive-figure/ socially-constructed-person	Y
Person/Conference_contributor/ Regular_author/ Contribution_co-author	null	-
Conference_document/Conference_ contribution/ Written_contribution/ Regular_contribution/ Extended_abstract	particular/spatio-temporal-particular/ endurant/non-physical-endurant/ non-physical-object/social-object/ non-agentive-social-object/ collection/ non-physical-collection	N
Person/ Track-workshop_chair	particular/spatio-temporal-particular/ endurant/ physical-endurant	N

Nossa abordagem apresenta resultados preliminares e a análise foi desenvolvida utilizando apenas uma ontologia de domínio, além disso, a ontologia utilizada não contém a descrição do conceito e isso pode ter gerado um contexto que não é rico o suficiente para desambiguar os *synsets* selecionados. Ainda assim, podemos observar resultados promissores com o uso do WordNet atuando como uma camada intermediária entre ontologias de topo e de domínio na tarefa de alinhar ontologias com diferentes níveis de abstração na representação dos conceitos.

5. Trabalhos Relacionados

Na literatura, diversas abordagens utilizam o WordNet na tarefa de alinhamento de ontologias de domínio para melhorar as métricas de similaridade. Em [Kwak and Yong 2010], a similaridade léxica para *matching* de ontologias considera a taxa de inclusão de nomes de entidades ontológicas e palavras de *synset* do WordNet em um conjunto agregado (incluindo relações de *hypernyms*, *hyponyms*, *holonyms* e *meronyms*). De maneira similar a nossa abordagem, os autores sobrepõe conjuntos de termos construídos a partir de informações disponíveis sobre entidades de ontologias. No entanto, em

Tabela 3. Fragmento das correspondências geradas entre a ontologia de domínio e a ontologia de topo SUMO

Conceitos de Conference	SUMO	Avaliação Manual
Person/Conference_participant/ Active_conference_participant	Entity/physical/object/ agent/IT_agent/ Human	Y
Person/Conference_contributor/ Invited_speaker	Entity/abstract/attribute/ relational_attribute/ SocialRole	Y
Person/Conference_contributor/ Regular_author/ Contribution_co-author	null	-
Conference_document/Conference_ contribution/ Written_contribution/ Regular_contribution/ Extended_abstract	Entity/ Abstract	N
Person/ Track-workshop_chair	Entity/physical/object/artifact/ furniture/seat/ Chair	N

nosso trabalho, exploramos o contexto das entidades ontológicas, incluindo seus sub-conceitos, super-conceitos e anotações, mas limitando-se a relação de hiperonímia. Em [Yatskevich and Giunchiglia 2004], é proposto um conjunto de *element-level matchers* que utilizam o WordNet. Os autores distinguem *knowledge-based matchers* (que exploram as propriedades estruturais do WordNet, frequentemente combinados com estatísticas de corpora em larga escala) e *gloss-based matchers* (que exploram a descrição dos *synsets* em WordNet). Nós também exploramos as propriedades estruturais do WordNet, mas com foco nas relações de hiperonímia e os *glosses* dos *synsets*. No entanto, em [Yatskevich and Giunchiglia 2004], não é explorada a noção de contexto de entidades ontológicas. Para medidas de similaridade terminológica, o *matcher* automático YAM++ [Ngo and Bellahsene 2013] inclui métricas baseadas em WordNet. O *matcher* XMAP [Djeddi and Khadir 2014] define uma medida de similaridade semântica usando o recurso UMLS (*Unified Medical Language System*) e WordNet. Em [Schmidt et al. 2016b], WordNet é usado para validar automaticamente conjuntos de correspondências gerados por *matchers*. Nos experimentos realizados, a validação envolve o alinhamento resultante entre uma ontologia de domínio e três ontologias de topo.

Em relação ao uso de ontologias de topo na tarefa de alinhamento de ontologias, [Lacasta et al. 2013] propõe um processo para converter automaticamente um modelo de conhecimento em uma ontologia através do alinhamento com a ontologia de topo DOLCE. Os autores têm o objetivo de transformar um tesouro e uma base de dados bibliográficos em uma ontologia. De acordo com os autores, o alinhamento prévio entre a ontologia DOLCE e WordNet foi utilizado como um passo intermediário para a identificação de correspondências entre as bases de dados e a ontologia de topo. As ontologias de topo são também usadas como um recurso para obter ou melhorar o alinhamento entre ontologias de domínio, como apresentado em [Padilha et al. 2012, Silva et al. 2011]. Em [Padilha et al. 2012], os autores alinham manualmente ontologias do domínio de conferência do OAEI utilizando a ontologia de topo UFO. O processo consiste em adotar um conjunto de padrões definidos pelos autores e fundamentados pela UFO. Assim, para cada conceito de domínio, é identificada uma correspondência em UFO, então, após as ontologias de domínio serem mapeadas para UFO, as mesmas são alinhadas entre si com base nos padrões especificados. Já em [Silva et al. 2011], o objetivo é auxiliar na identificação

de termos equivalentes entre a GO (*genome ontology*) e outras ontologias biomédicas. A abordagem consiste em uma fase de preparação que identifica a ontologia de origem, destino e de topo que será utilizada no processo. A abordagem é semi-automática, onde o alinhamento entre os conceitos de domínio e topo ocorre manualmente, após, o alinhamento entre ontologias de mesmo domínio é realizada de forma automática por meio de um algoritmo que implementa métricas de similaridade extensional, similaridade de domínio e escopo e similaridade de conceitos. A ontologia de topo BFO é adotada nos experimentos.

O nosso trabalho também foca no alinhamento entre ontologias de topo e de domínio, mas diferentemente de [Padilha et al. 2012, Silva et al. 2011], o alinhamento entre as ontologias é realizado automaticamente e utiliza o WordNet como uma camada intermediária, visto que existem alinhamentos prévios do WordNet com as ontologias de topo utilizadas. Além do mais, a nossa abordagem adota a noção de contexto que é construída a partir de informações extraídas das ontologias, e não é explorada nos trabalhos mencionados.

6. Conclusão e Trabalhos Futuros

Alinhar ontologias de domínio e topo é um desafio, em particular dado os diferentes níveis de abstração destas ontologias. Este trabalho apresenta uma abordagem inicial para o alinhamento entre ontologias de topo e domínio, que utiliza o WordNet como uma camada intermediária. Foi realizada uma avaliação preliminar, cujos resultados apontam para as vantagens e limitações da nossa abordagem. Em particular, observamos que alguns contextos não são ricos o suficiente para realizar a desambiguação; e o problema da recuperação de *synsets* nulos, quando não foram identificados os termos no WordNet. Ainda assim, consideramos que o uso do WordNet atuando como uma camada intermediária entre ontologias de topo e de domínio é uma abordagem promissora. Nos experimentos realizados, observamos que a maioria dos conceitos representados na ontologia de domínio participaram de alinhamentos com a ontologia de topo, por meio do WordNet. Além do mais, observamos um número razoável de correspondências consideradas corretas na avaliação manual, fato que não ocorre quando utilizamos diretamente ferramentas de matching de ontologias disponíveis (como abordado em [Schmidt et al. 2016a]). No entanto, estamos cientes de que, no caso de domínios mais específicos, pode ocorrer a falta de cobertura no WordNet. Dessa forma essa abordagem é restrita aos casos de domínios mais gerais como o caso do domínio Conferência, estudado neste trabalho.

Como trabalhos futuros, pretendemos explorar outras técnicas de alinhamento de ontologias com o objetivo de apoiar a identificação de correspondências entre conceitos, principalmente a relação de subsunção que é altamente desejável quando alhamos ontologias de domínio e topo. Pretendemos também desenvolver experimentos utilizando outras ontologias de domínio para que possamos obter uma avaliação mais abrangente da nossa abordagem.

Referências

- Djeddi, W. E. and Khadir, M. T. (2014). A novel approach using context-based measure for matching large scale ontologies. In *Data Warehousing and Knowl. Discovery*, pages 320–331.

- Euzenat, J. and Shvaiko, P. (2007). *Ontology Matching*. Springer-Verlag, 1st edition.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). Sweetening Ontologies with DOLCE. In *Knowl. Engin. and Knowl. Manag.*, pages 166–181.
- Kwak, J. and Yong, H. (2010). Ontology matching based on hypernym, hyponym, homonym and meronym sets in wordnet. *Web and Semantic Technology*, 1(2):1–14.
- Lacasta, J., Nogueras-Iso, J., Falquet, G., Teller, J., and Zarazaga-Soria, F. J. (2013). Design and Evaluation of a Semantic Enrichment Process for Bibliographic Databases. *Data Knowledge Engineering*, 88:94–107.
- Maedche, A. and Staab, S. (2002). Measuring similarity between ontologies. In *Knowledge Engineering and Knowledge Management*, pages 251–263.
- Mascardi, V., Cordì, V., and Rosso, P. (2007). A Comparison of Upper Ontologies. In *Joint Workshop on Agents and Industry*, pages 55–64.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications ACM*, 38(11):39–41.
- Ngo, D. and Bellahsene, Z. (2013). YAM++ - Results for OAEI 2013. In *Workshop on Ontology Matching*, pages 211–218.
- Niles, I. and Pease, A. (2001). Towards a Standard Upper Ontology. In *Formal Ontology in Information Systems*, pages 2–9.
- Niles, I. and Pease, A. (2003). Mapping wordnet to the sumo ontology. In *Knowledge Engineering Conference*, pages 23–26.
- Padilha, N. F., Baião, F., and Revoredo, K. (2012). Alignment Patterns based on Unified Foundational Ontology. In *Brazilian Ontology Research Seminar*, pages 48–59.
- Schadd, F. C. and Roos, N. (2012). Coupling of wordnet entries for ontology mapping using virtual documents. In *Workshop on Ontology Matching*, pages 25–36.
- Schmidt, D., Trojahn, C., and Vieira, R. (2016a). Analysing Top-level and Domain Ontology Alignments from Matching Systems. In *Workshop on Ontology Matching*, pages 1–12.
- Schmidt, D., Trojahn, C., Vieira, R., and Kamel, M. (2016b). Validating Top-level and Domain Ontology Alignments using WordNet. In *Brazilian Ontology Research Seminar*, pages 119–130.
- Silva, V. S., Campos, M. L. M., Silva, J. C. P., and Cavalcanti, M. C. (2011). An approach for the alignment of biomedical ontologies based on foundational ontologies. *Information and Data Management*, 2(3):557–572.
- Wang, P. (2011). Lily Results on SEALS Platform for OAEI 2011. In *Workshop on Ontology Matching*, pages 156–162.
- Yatskevich, M. and Giunchiglia, F. (2004). Element level semantic matching using WordNet. In *Meaning Coordination and Negotiation Workshop*, pages 37–48.