

Uma análise da avaliação dos resultados da medida de Similaridade Semântica (SiSe) entre ontologias em português

Juliano Baldez de Freitas, Vera Lúcia Strube de Lima

Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Programa de Pós-Graduação em Ciência da Computação - PPGCC
Avenida Ipiranga, 6681 - Prédio 32 - Partenon
CEP 90619-900 Porto Alegre - RS - Brasil

{jfreitas, vera}@inf.pucrs.br

Abstract. *This article concerns the analysis of the evaluation results of the SiSe measure (Semantic Similarity), that is used in the mapping between ontologies. SiSe, which is an extension of the proposal of Maedche and Staab [Maedche and Staab 2002], calculates the similarity between Portuguese ontologies using Natural Language Processing techniques. This measure helps the semi-automatic mapping between Portuguese ontologies aiming at information reuse and integration.*

Resumo. *Este artigo relata a avaliação dos resultados da medida SiSe (Similaridade Semântica) utilizada no mapeamento entre ontologias. SiSe calcula a similaridade entre ontologias em Português adaptando a proposta denominada Mapeamento Taxonômico (MT) de Maedche e Staab [Maedche and Staab 2002]. Para o cálculo da similaridade, SiSe utiliza técnicas de Processamento da Linguagem Natural. Esta medida é utilizada para auxiliar o mapeamento semi-automático entre ontologias em português, visando o reuso e a integração de informação.*

1. Introdução

Atualmente o uso e reuso do conhecimento são fundamentais, devido à crescente quantidade de informação que está sendo gerada. Este fato faz com que pessoas e organizações tenham de gerenciar seu conhecimento de modo mais eficaz. Combinar conhecimentos de domínios distintos pode acarretar problemas como, por exemplo, formatos de representação do conhecimento distintos e inconsistências semânticas, entre outros [Ding and Foo 2002]. Segundo Chaves em [Chaves 2003], inconsistências semânticas podem ser geradas quando dois sistemas projetados independentemente são integrados, ocorrendo alguns mal entendidos entre os significados e as interpretações para o mesmo dado como, por exemplo, nomes, estruturas, esquemas ou atributos, entre outros. Estas inconsistências semânticas também se aplicam à área de Engenharia Ontológica.

A integração ou determinação de equivalência das informações entre duas estruturas ontológicas¹ depende do mapeamento realizado entre os termos destas duas ontologias. Para Maedche e Staab em [Maedche and Staab 2002], mapear termos de duas ou mais ontologias é associar conceitos equivalentes, de acordo com uma relação de similaridade.

¹Em nosso trabalho, os termos ontologia e estrutura ontológica são usadas como sinônimos

Para evitarmos as inconsistências semânticas obtidas ao integrar e reusar informações entre ontologias construídas individualmente, o mapeamento entre as mesmas ocorre através da utilização de medidas de similaridade que comparam os elementos destas estruturas e identificam suas semelhanças, evitando que estas inconsistências ocorram. Atualmente as pesquisas realizadas mostram um grande esforço para automatizar o processo de mapeamento entre ontologias, dado que o mapeamento manual entre ontologias não é escalável em boa parte dos casos (por exemplo, cita-se o contexto da Web).

Este artigo tem como objetivo apresentar e discutir os resultados obtidos na avaliação da medida SiSe, a qual calcula a similaridade entre ontologias em Português. O artigo está organizado em 5 seções. A Seção 2 apresenta os conceitos sobre o mapeamento entre ontologias, bem como uma breve descrição das medidas de similaridade lexical e semântica descritas em trabalhos correlatos. A Seção 3 descreve a medida de similaridade semântica SiSe que adapta a abordagem MT de Maedche e Staab. Na Seção 4, são apresentados o método de avaliação e os resultados obtidos. A Seção 5 traz algumas considerações, contribuições e sugestões de trabalhos futuros em relação à medida SiSe.

2. Mapeamento entre ontologias

Gruber, em [Gruber 1995], define uma ontologia como “*uma especificação explícita de uma conceitualização*”. Nesta definição, *conceitualização* diz respeito a um modelo abstrato de um determinado fenômeno do mundo, usualmente restrito ao domínio deste fenômeno; o termo *explícita* indica que os conceitos e relações do modelo abstrato são explícitos em termos e definições.

O mapeamento entre ontologias tem como finalidade o uso e reuso de ontologias, expansão e combinação das mesmas, com o intuito de aumentar a informação e conhecimento em diferentes domínios que são integrados para suportar nova comunicação e uso [Ding and Foo 2002]. A Figura 1 ilustra o processo de mapeamento entre ontologias onde duas ontologias, 1 e 2, tem seus elementos mapeados (linhas pontilhadas) através de uma medida de similaridade.

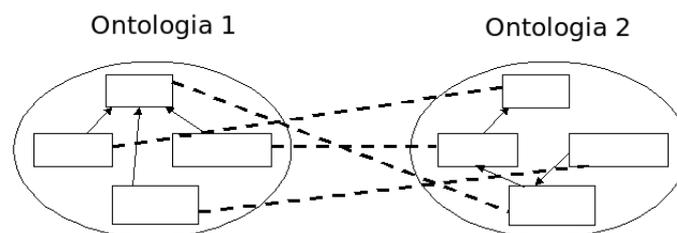


Figura 1. Mapeamento entre ontologias

No mapeamento entre ontologias construídas individualmente, termos com mesmo significado podem não ser considerados similares pelas medidas de similaridade mais usuais. Isto se deve a dois fatores importantes:

- **Linguagem Natural:** de acordo com Maedche e Staab [Maedche and Staab 2002], o inconveniente é que as ontologias reais que conhecemos não especificam suas conceitualizações somente por estruturas lógicas, mas por uma referência de termos fundamentada em linguagem natural;

- **Estrutura Taxonômica:** ontologias também podem conter más combinações em nível de modelo estrutural como, por exemplo, taxonomias distintas [Noy 2004].

A construção de ferramentas e técnicas que permitam o mapeamento automático ou semi-automático entre ontologias é importante área de pesquisa atualmente. Segundo Noy [Noy 2004], para permitir o mapeamento de forma automática ou semi-automática, estas ferramentas devem analisar as seguintes características na definição das ontologias envolvidas no processo: *nomes dos conceitos e descrições em linguagem natural; hierarquia das classes (relacionamento de subclasses e superclasses); definições de propriedades (domínio, abrangência, restrições); instâncias das classes; descrições das classes.*

Na literatura são encontradas várias abordagens para medir a similaridade entre elementos de ontologias. Classificamos estas abordagens em dois grupos de medidas: similaridade lexical e similaridade semântica, embora se observe que esta classificação não distingue exatamente os dois grupos.

Na similaridade lexical podemos destacar alguns trabalhos como, por exemplo: Combinação de Caracteres (CC) [Maedche and Staab 2002], Similaridade Lexical (SL) [Chaves 2003] e FCA-Merge [Stumme and Maedche 2001]. Nestas abordagens normalmente são usadas soluções que medem a similaridade entre as cadeias de caracteres, bem como fazem uso de heurísticas. A medida de similaridade normalmente é dada através de coeficientes, os quais medem lexicalmente a proximidade dos elementos.

A similaridade semântica aborda uma comparação ligada ao significado da expressão utilizada por um termo na ontologia, bem como as relações semânticas que este termo possui na ontologia. Nesta abordagem podemos destacar os seguintes trabalhos correlatos: Mapeamento Taxonômico (MT) [Maedche and Staab 2002]; Combinação Semântica (CS) [Giunchiglia and Shvaiko 2004]; Cupid [Madhavan et al. 2001]; Mafra [Maedche et al. 2002]; e Prompt [Noy and Musen 2003]. Alguns destes trabalhos correlatos não são diretamente aplicados a nossa língua, devido à carência de recursos de Processamento da Linguagem Natural (por exemplo, bases de dados lexicais, Tesouro de sinônimos, etc.) para o português, dos quais estes trabalhos correlatos fazem uso em suas propostas.

Desta forma, a abordagem semântica mais natural a ser adaptada para medir a similaridade entre ontologias em português é o Mapeamento Taxonômico (MT), de Maedche e Staab [Maedche and Staab 2002], que não necessita recursos adicionais de processamento da língua para o cálculo da similaridade. Em nosso entendimento, a utilização de recursos de PLN disponíveis para o português pode gerar ganhos, em alguns casos, em relação à medida original MT.

3. Medida de Similaridade Semântica (SiSe)

Nossa contribuição para medir a similaridade semântica entre ontologias em português é feita através da abordagem denominada **Similaridade Semântica (SiSe)**, descrita em detalhe em [Freitas 2007].

Para definirmos a medida SiSe realizamos algumas adaptações na medida MT no que se refere aos conceitos do Semantic Cotopy (SC) [Maedche and Staab 2002] e Common Semantic Cotopy (CSC) [Cimiano et al. 2005]. Estas adaptações utilizam um

recurso de PLN, um algoritmo de *stemming*², e foram denominadas SC' e CSC' . O uso do *stemming* prioriza o nível lexical no intuito de encontrar termos lexicalmente similares para a comparação semântico-estrutural.

A medida MT adota uma abordagem que compara a hierarquia das ontologias. Ontologias criadas por diferentes especialistas podem diferir na representação hierárquica para um mesmo conceito, ou seja, cada especialista tem uma visão diferente de um determinado domínio, e estas diferenças são visíveis através da construção de hierarquias distintas. Este fato faz com que a medida MT, que é baseada nos superconceitos e subconceitos dos termos, possa mascarar ou ressaltar algumas similaridades entre os termos das ontologias. Termos que são semanticamente similares podem estar dispostos na hierarquia de tal forma que seus superconceitos e subconceitos sejam diferentes, fazendo com que a medida MT retorne um coeficiente de similaridade baixo. No entanto, esta similaridade hierárquica pode ressaltar similaridades tais como a existência de termos com representações lexicais distintas mas que, no entanto, possuem superconceitos e subconceitos similares. Isto pode indicar que estes termos são similares semanticamente. A seguir apresentamos a medida SiSe e suas abordagens (SC' e CSC') para comparação das hierarquias entre dois termos de ontologias distintas.

3.1. Semantic Cotopy (SC')

Primeiramente modificamos a equação original do SC [Maedche and Staab 2002], chegando às adaptações SC' e SiSe representadas pelas equações 1 e 2.

$$SC'(c_i, O_1) := \{\Delta_{c_j} \in C_i \mid c_i \leq_{C_1} c_j \text{ ou } c_j \leq_{C_1} c_i\} \quad (1)$$

$$SiSe(c_1, O_1, c_2, O_2) = \frac{|SC'(c_1, O_1) \cap SC'(c_2, O_2)|}{|SC'(c_1, O_1) \cup SC'(c_2, O_2)|} \in [0, 1] \quad (2)$$

Na Equação 1, o símbolo Δ representa o *stem* do conceito c_j , possibilitando que as variações lexicais dos termos possam ser consideradas através de uma única representação, e permitindo que termos similares tenham um coeficiente de similaridade semântica maior, na comparação das hierarquias. Estes conjuntos representados pelos *stems* dos termos são comparados através da medida de Jaccard [Manning and Schütze 1999] (assim como na medida MT) de acordo com a Equação 2.

Os trechos de ontologias³ utilizados como exemplo de hierarquias estão representados na Tabela 1. O SC' de cada termo das ontologias é representado por seu *stem*. Como exemplo, comparamos a similaridade dos termos **partido político** em O_1 e **partidos políticos** em O_2 de acordo com SC' . Em O_1 **partido político** tem como superconceitos os termos **direito eleitoral** e **direito constitucional**, e não possui subconceitos. O conjunto formado pelo SC' é representado pelo *stem* de cada um destes termos, **partPolitic**, **direitConstituc** e **direitEleitor**, respectivamente, formando o conjunto $\{\text{partPolitic}, \text{direitConstituc}, \text{direitEleitor}\}$.

²O algoritmo de *stemming* utilizado é o PortugueseStemmer [Orengo and Huyck 2001]

³Estes trechos de ontologias foram extraídos do Vocabulário Controlado Básico do Senado (O_1) e Vocabulário Controlado da USP (O_2)

direitConstituc, direitEleitor}. O termo partidos políticos em O_2 não possui subconceitos e tem como superconceitos os termos direito eleitoral e direito. Desta forma o conjunto resultante, de acordo com o SC' é {partPolitic, direitEleitor, direit}.

Tabela 1. Trechos de hierarquias do domínio do Direito extraídas de duas ontologias

Ontologia 1 (O_1)	Ontologia 2 (O_2)
01 direito constitucional	01 direito
02 direito eleitoral	02 direito eleitoral
03 campanha eleitoral	03 crime eleitoral
04 eleição	04 domicílio eleitoral
05 partido político	05 eleições
06 sistema eleitoral	06 justiça eleitoral
07 voto	07 partidos políticos
	08 sistema distrital
	09 voto

Aplicamos a medida de Jaccard para comparar a similaridade semântica dos conjuntos formados pelo SC' dos termos das ontologias O_1 e O_2 , de acordo com a Equação 2. Como coeficiente de similaridade entre os dois termos resulta o valor 0.5, de acordo com a SiSe. Este valor é considerado baixo, e isto pode ser explicado pelo fato de a medida comparar a estrutura hierárquica dos termos, ocorrendo que um termo pode ter um maior número de superconceitos e subconceitos do que o outro termo na outra ontologia, baixando o coeficiente de similaridade. A seguir observamos as etapas do cálculo da similaridade entre estes dois termos:

$$SiSe = \frac{|SC'(\text{partido político}, O_1) \cap SC'(\text{partidos políticos}, O_2)|}{|SC'(\text{partido político}, O_1) \cup SC'(\text{partidos políticos}, O_2)|}$$

$$SiSe = \frac{|\{\text{direitConstituc, direitEleitor, partPolitic}\} \cap \{\text{direit, direitEleitor, partPolitic}\}|}{|\{\text{direitConstituc, direitEleitor, partPolitic}\} \cup \{\text{direit, direitEleitor, partPolitic}\}|}$$

$$SiSe = \frac{|\{\text{direitEleitor, partPolitic}\}|}{|\{\text{direit, direitConstituc, direitEleitor, partPolitic}\}|}$$

$$SiSe = \frac{2}{4} \in [0, 1]$$

3.2. Common Semantic Cotopy (CSC')

Apresentamos, também, uma adaptação na medida MT no que se refere ao conceito do CSC [Cimiano et al. 2005]. Modificamos a definição original do CSC , e apresentamos o CSC' e SiSe, de acordo com as Equações 3 e 4.

$$CSC'(c_i, O_1, O_2) := \{\Delta_{c_j} \in \Delta_{C_1} \cap \Delta_{C_2} | c_i \leq_{C_1} c_j \text{ ou } c_j \leq_{C_1} c_i\} \quad (3)$$

$$SiSe(c_1, O_1, c_2, O_2) = \frac{|CSC'(c_1, O_1, O_2) \cap CSC'(c_2, O_2, O_1)|}{|CSC'(c_1, O_1, O_2) \cup CSC'(c_2, O_2, O_1)|} \in [0, 1] \quad (4)$$

Na Equação 3, o símbolo Δ representa o *stem* do termo c_j em questão. O conjunto CSC' associado a um termo é formado com base em seus subconceitos e superconceitos que são comuns em ambas as ontologias. Estes termos comuns são representados através de seus *stems* Δ_{C_1} e Δ_{C_2} , formando um conjunto $\Delta_{C_1} \cap \Delta_{C_2}$. Desta forma, o *stem* de um subconceito ou superconceito Δ_{c_j} fará parte do conjunto CSC' se o mesmo aparecer em ambas as hierarquias. Os conjuntos dos termos de cada ontologia formados pelo CSC' são comparados através da medida de Jaccard, conforme a Equação 4.

Para exemplo de cálculo da similaridade utilizamos as ontologias representadas nas hierarquias da Tabela 1. O conjunto de termos comuns para estas ontologias, $\Delta_{C_1} \cap \Delta_{C_2}$, é {partPolitic, vot, direitEleitor, ele}.

Por exemplo, ao compararmos a similaridade entre os termos partido político em O_1 e partidos políticos em O_2 , temos como CSC' do termo partido político seus superconceitos direito eleitoral e direito constitucional, não contendo nenhum subconceito. Através do algoritmo de *stemming* temos os respectivos *stems* para os termos: partPolitic, direitEleitor, direitConstituc. Analisando os termos comuns em ambas as ontologias o termo partido político em O_1 tem como CSC' o conjunto {partPolitic, direitEleitor}, pois direitConstituc $\notin \Delta_{C_1} \cap \Delta_{C_2}$.

O termo partidos políticos em O_2 também não possui subconceitos, e tem como superconceitos os termos direito eleitoral e direito. Os termos são representados pelos *stems* partPolitic, direitEleitor e direit. Verifica-se quais destes *stems* são comuns às duas ontologias e tem-se o conjunto CSC' formado por {partPolitic, direitEleitor}. O termo direit não faz parte do conjunto pois não é um termo comum às ontologias.

Dados os dois conjuntos para cada termo, de acordo com o CSC' , se aplica a medida de Jaccard para medir a similaridade semântica (Equação 4). A seguir apresentamos a sequência do cálculo de similaridade entre estes dois termos.

$$SiSe = \frac{|CSC'(\text{partido político}, O_1, O_2) \cap CSC'(\text{partidos políticos}, O_2, O_1)|}{|CSC'(\text{partido político}, O_1, O_2) \cup CSC'(\text{partidos políticos}, O_2, O_1)|}$$

$$SiSe = \frac{|\{\text{direitEleitor}, \text{partPolitic}\} \cap \{\text{direitEleitor}, \text{partPolitic}\}|}{|\{\text{direitEleitor}, \text{partPolitic}\} \cup \{\text{direitEleitor}, \text{partPolitic}\}|}$$

$$SiSe = \frac{|\{\text{direitEleitor}, \text{partPolitic}\}|}{|\{\text{direitEleitor}, \text{partPolitic}\}|}$$

$$SiSe = \frac{2}{2} \in [0, 1]$$

4. Avaliação

Lin [Lin 1998] relata que não existe um modo padrão para avaliação de medidas de similaridade, e que se trata de uma tarefa muito subjetiva.

Analisando as iniciativas de avaliação dos trabalhos correlatos, notamos que a comparação com outras abordagens seria difícil devido ao fato de as mesmas serem desenvolvidas e avaliadas para ontologias em outras línguas (por exemplo, inglês e alemão) com diferentes regras de formação lexical e morfológica. Desta maneira o mais natural para proceder à avaliação foi recorrer à comparação de resultados entre os coeficientes das medidas SiSe e MT, com a análise humana.

Para tal, empregamos inicialmente uma metodologia que produziu a um “*Golden Mapping*” ou “Mapeamento Dourado” entre as ontologias comparadas. As ontologias utilizadas no processo de avaliação da medida SiSe foram o Vocabulário Controlado Básico do Senado Federal⁴ e o Vocabulário Controlado da USP⁵. Foram escolhidos 5 pares de ontologias do domínio do Direito e submetidos à avaliação humana, descrita na Seção 4.1. A Tabela 2 apresenta o número de termos em cada par de hierarquias considerado.

Tabela 2. Informações sobre os pares de ontologias utilizados na avaliação

Ontologias	Número de termos		Total de termos
	VCUSP	VCBS	
Par 1	21	23	44
Par 2	17	16	33
Par 3	20	16	36
Par 4	40	34	74
Par 5	28	19	47
Total	126	108	234

4.1. *Golden Mapping*

O “*Golden Mapping*” (GM) consiste em realizar a avaliação humana antes da obtenção dos resultados da medida SiSe. Este fato faz com que a avaliação seja menos tendenciosa, pois os humanos podem sugerir mapeamentos que as medidas não encontram. O GM constitui um consenso dos mapeamentos de acordo com as avaliações humanas, podendo ser confrontado com os resultados das medidas automáticas para o mapeamento entre ontologias.

Estabelecemos, para este trabalho, uma avaliação humana feita por três indivíduos com especialidades distintas. São eles:

1. Lingüista;
2. Bacharel em Ciência da Computação;
3. Bacharel em Direito.

A escolha das especialidades mencionadas se baseia no conhecimento das relações e significados dos termos das ontologias pelo diferentes profissionais especialistas, sendo dada maior ponderação ao terceiro.

A metodologia de nossa avaliação consiste em quatro passos. Ao final destes passos têm-se uma referência para confrontar com o resultado das medidas de similaridade, em nosso caso com as medidas SiSe e MT. A seguir temos a descrição destes passos:

⁴Pode ser acessado *on-line* em <http://webthes.senado.gov.br/thes/default-vcbs.htm>

⁵Disponível para acesso em <http://143.107.73.99/Vocab/SIBIX652.dll/Index>

- **pares de ontologias:** cada um dos humanos envolvidos no processo de avaliação recebe um documento contendo: (i) a estrutura hierárquica dos trechos de ontologias; (ii) uma tabela de mapeamento para cada par de hierarquias (extratos) de ontologias.
- **avaliação humana:** cada avaliador humano indica, em uma tabela de mapeamento, os termos que foram considerados similares, assinalando a que ontologia os termos pertencem. Desta forma, os avaliadores preenchem uma tabela de mapeamento para cada par de extratos de ontologias.
- **análise da avaliação humana:** após realizada a avaliação buscamos um consenso dos termos mapeados pelos participantes, de acordo com as seguintes regras: será considerado para o GM aquele mapeamento entre os termos das ontologias que foi identificado por no mínimo dois dos três humanos. Os mapeamentos assinalados pelo Bacharel em Direito são sempre considerados, independente da análise dos outros humanos, devido ao seu conhecimento específico dos termos do domínio que lhe permite encontrar mapeamentos que as medidas automáticas e os outros humanos não detectam.
- **Golden Mapping:** ao final deste processo é criada uma referência de mapeamentos entre as ontologias envolvidas. Esta referência, denominada “Golden Mapping”, é confrontada com o mapeamento das medidas de similaridade.

4.2. Resultados obtidos

Analisando os resultados obtidos com as medidas SiSe e MT, comparados ao GM, observamos que a medida que em geral encontrou o maior número de mapeamentos para cada par de ontologias foi a medida SiSe utilizando o CSC' . Como podemos ver na Tabela 3, no Par 5 esta medida detectou 100% dos mapeamentos do GM.

Tabela 3. Resumo dos resultados obtidos com as medidas MT e SiSe em relação ao Golden Mapping (GM)

	GM	MT		SiSe	
		SC	CSC	SC'	CSC'
Par 1	14	11 (78.57%)	12 (85.71%)	11 (78.57%)	12 (78.57%)
Par 2	11	6 (54.54%)	8 (72.72%)	6 (54.54%)	8 (72.72%)
Par 3	13	7 (53.84%)	9 (69.23%)	7 (53.84%)	9 (69.23%)
Par 4	21	12 (57.12%)	20 (95.23%)	16 (76.16%)	21 (100%)
Par 5	15	4 (26.66%)	7 (46.66%)	4 (26.66%)	7 (46.66%)

Para os pares de hierarquias similares, o número de mapeamentos não variou muito entre as medidas, como foi o caso dos Pares 1, 2 e 3. Também percebemos que, quando as ontologias possuem termos com pequenas diferenças lexicais, as abordagens da medida SiSe que utilizam o *stemming* (SC' e CSC') obtêm melhores resultados. O CSC e CSC' encontram um maior número de mapeamentos quando as hierarquias das ontologias possuem níveis diferentes, em relação ao SC e CSC . No entanto, também encontram um número elevado de falsos positivos (termos que não são semanticamente similares de acordo com o GM gerado para os pares de ontologias). Por exemplo, para o Par 4 foram encontrados 10 falsos positivos utilizando o CSC , e 26 falsos positivos utilizando o CSC' .

Apesar de a medida SiSe não ter obtido um ganho muito significativo em relação à medida MT, a mesma produz melhores resultados quando os termos das ontologias são representados com variações léxico-morfológicas, permitindo que a utilização do *stemming* resulte num ganho para o cálculo da similaridade entre ontologias.

Dentre as limitações deste trabalho, pode-se incluir a carência de dados como ontologias em português descritas em linguagens de marcação semântica (por exemplo, RDFS e OWL). Isto faz com que o desenvolvimento de medidas de similaridade para nossa língua fique limitado à análise dos rótulos (cadeias de caracteres) dos termos (classes) e das relações de hierarquia, não permitindo, por exemplo, a análise das definições de propriedades (domínio, abrangência e restrições) e instâncias, entre outros.

5. Considerações finais

Uma lacuna da área, que pretendemos preencher com este trabalho, é a investigação no que se refere ao mapeamento entre ontologias na língua portuguesa. Poucas abordagens testadas ou propostas para nossa língua foram encontradas. Percebeu-se como causa para a falta de abordagens semânticas para ontologias na língua portuguesa, o fato de a língua portuguesa não contar ainda com ferramentas e recursos lingüísticos fundamentais para o PLN, como uma base de dados lexical consistente que possa identificar as relações semânticas e significados das palavras da língua. Mas isto não impede que alternativas sejam pesquisadas, fazendo uso das ferramentas existentes.

A nosso ver, as principais contribuições deste trabalho foram: (i) a adaptação de uma medida de similaridade semântica para ontologias em português, com experimentação e avaliação dos resultados, através da medida SiSe; (ii) a avaliação, com o uso de um *Golden Mapping*.

Diante dos estudos realizados e dos resultados obtidos podemos ver como continuidade os seguintes trabalhos: (i) comparação da medida SiSe com outras medidas de similaridade para o português; (ii) aplicação da medida SiSe a outros idiomas, utilizando o algoritmo de *stemming* específico para a língua utilizada; (iii) utilização de uma base de dados lexical ou dicionário de sinônimos para encontrar termos com mesmo significado mas com representações lexicais distintas, aumentando o número de mapeamentos entre as ontologias; (iv) enriquecimento das heurísticas, permitindo a identificação de um maior número de mapeamentos.

Referências

- Chaves, M. S. (2003). Mapeamento e comparação de similaridade entre estruturas ontológicas. Mestrado, Programa de Pós-Graduação em Ciência da Computação, PU-CRS. 93 p.
- Cimiano, P., Hotho, A., and Staab, S. (2005). Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research - JAIR*, 24:263–303.
- Ding, Y. and Foo, S. (2002). Ontology research and development. Part 2 - a review of ontology mapping and evolving. *Journal of Information Science*, 28(5):375–388.

- Freitas, J. B. (2007). Sise: Medida de similaridade semântica entre ontologia em português. Mestrado, Programa de Pós-Graduação em Ciência da Computação, PUCRS. 106 p.
- Giunchiglia, F. and Shvaiko, P. (2004). Semantic matching. *The Knowledge Engineering Review*, 18(3):265–280.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. In: *International Journal of Human-Computer Studies*, 43(5/6):907–928.
- Lin, D. (1998). An information-theoretic definition of similarity. In: *Proceedings of the 15th International Conference on Machine Learning*, pages 296–304.
- Madhavan, J., Bernstein, P. A., and Rahm, E. (2001). Generic schema matching using Cupid. In: *Proceedings of the 27th Very Large Data Bases (VLDB)*, pages 48–58.
- Maedche, A., Motik, B., Silva, N., and Volz, R. (2002). MAFRA - A MAPPING FRAMEWORK for distributed ontologies. In: *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*.
- Maedche, A. and Staab, S. (2002). Measuring similarity between ontologies. In: *Proceedings of the European Conference on Knowledge Acquisition and Management (EKAW)*.
- Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, Massachusetts, 1st edition.
- Noy, N. F. (2004). Semantic integration: a survey of ontology-based approaches. *SIGMOD Record*, 33(4):65–70.
- Noy, N. F. and Musen, M. A. (2003). The PROMPT suite: Interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, 59(6):983–1024.
- Orengo, V. M. and Huyck, C. (2001). A stemming algorithm for the Portuguese language. In: *Proceedings of the 8th International Symposium on String Processing and Information Retrieval (SPIRE-2001)*, pages 186–193.
- Stumme, G. and Maedche, A. (2001). FCA-Merge: bottom-up merging of ontologies. In: *Proceedings of the 17th Intl. Conf. on Artificial Intelligence (IJCAI '01)*, pages 225–230.