

O papel das relações semânticas em português:

Comparando o TeP, o MWN.PT e o o PAPEL

Diana Santos, Anabela Barreiro, Luís Costa, Cláudia Freitas, Paulo Gomes, Hugo Gonçalo Oliveira, José Carlos Medeiros & Rosário Silva

Este artigo versa a descoberta e representação de uma língua através de uma ontologia lexical, assim como a aplicação prática das mesmas.

Iniciamos o artigo por uma discussão do interesse em compilar e tornar acessível um conjunto de relações semânticas entre palavras em português, semelhante ao feito para o inglês pela WordNet (Miller 95) e pela MindNet (Richardson et al 98).

Começamos por fazer uma panorâmica dos recursos semânticos de cobertura larga para o português, em particular dos projectos que conhecemos com o objectivo de tornar esse tipo de recursos acessíveis à comunidade em geral, o TeP (Dias da Silva & Moraes, 2003, Dias da Silva et al. 2000), o MWN.PT¹ e o PAPEL (Gonçalo Oliveira et al., 2008).

Apresentamos brevemente o nosso trabalho como uma tentativa de apressar semiautomaticamente esse processo, incorporando num ciclo comum que permita a vários grupos distintos, mas interessados numa descrição larga do português, juntar esforços e obter uma síntese de informação utilizável em aplicações de processamento computacional da nossa língua.

Fazemos depois o que pensamos ser a primeira comparação entre os vários recursos existentes, de duas formas:

- Quantitativa, seguindo as várias metodologias de que temos conhecimento na literatura sobre avaliação de ontologias
- Qualitativa, seleccionando um conjunto de entradas existentes nos três recursos e comparando a sua informação, no âmbito de quatro domínios semânticos: a cor, a causalidade, a localização espacial e o tempo e aspecto

Finalmente apresentamos algumas ideias e exemplos de como tais recursos podem ser utilizados em várias áreas, desde estudos exploratórios em linguística, avaliação de (outros)

¹ Este recurso, ao contrário dos outros, não é grátis, mas compra-se através da ELRA, http://catalog.elra.info/product_info.php?products_id=1101

recursos lexicais, criação de material didático e entretenimento educativo, recolha de informação e recolha de informação cruzada, resposta automática a perguntas, até paráfrase automática, ajuda à redacção e tradução automática.

Referências

- Dias-da-Silva, Bento Carlos & Helio Roberto de Moraes. "A construção de um thesaurus eletrônico para o português do Brasil". *ALFA* **47**, N. 2, 2003, pp. 101-115.
- Dias-da-Silva, Bento Carlos, Helio Roberto de Moraes, Mirna Fernanda de Oliveira, Ricardo Hasegawa, Daniela Amorim, Christie Passchoalino & Ana Cláudia Nascimento. "Construção de um thesaurus eletrônico para o português do Brasil". In Maria das Graças Volpe Nunes (ed.), *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)*, pp. 1-10.
- Gonçalo Oliveira, Hugo, Diana Santos, Paulo Gomes & Nuno Seco. "PAPEL: a dictionary-based lexical ontology for Portuguese". In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (eds.), *Computational Processing of the Portuguese Language, 8th International Conference, Proceedings (PROPOR 2008)* (Aveiro, Portugal, 8-10 de Setembro, 2008), Springer Verlag, pp. 31-40.
- Miller, George A. "WordNet: a lexical database for English". *Communications ACM* **38**, 11 (Nov. 1995), pp. 39-41.
- Richardson, Stephen, William B. Dolan & Lucy Vanderwende. "MindNet: acquiring and structuring semantic information from text". *Proceedings of the 17th International Conference on Computational Linguistics, COLING-ACL'98* (August 10-14, Montréal, Québec, Canadá), Vol. 2, 1998, pp. 1098-1102.