

**Referência**

FELIPPO, Ariani di. DELIMITAÇÃO DO DOMÍNIO DA EDUCAÇÃO A DISTÂNCIA COM VISTAS À CONSTRUÇÃO DA BASE LEXICAL WORDNET.EAD. In: SEMINÁRIO DO GEL, 58., 2010, *Programação...* São Carlos (SP): GEL, 2010. Disponível em: <<http://www.gel.org.br/?resumo=6632-10>>. Acesso em: dd.mmm.aaaa.

Dados do trabalho

Título: DELIMITAÇÃO DO DOMÍNIO DA EDUCAÇÃO A DISTÂNCIA COM VISTAS À CONSTRUÇÃO DA BASE LEXICAL WORDNET.EAD

Autor(es): ARIANI DI FELIPPO - UFSCAR

Resumo: No domínio do Processamento Automático das Línguas Naturais (PLN), o desenvolvimento de sistemas que processam língua natural (p.ex.: sistemas de tradução automática) pode requerer a utilização de certos recursos linguísticos, os chamados lingwares. Tais recursos desempenham papel central na arquitetura de um sistema de PLN. Esse é o caso, por exemplo, das "bases de conhecimento lexical". Dada a necessidade crescente de se processar textos especializados, bases de conhecimento lexical especializado (ou terminológico) passaram a ser desenvolvidas para várias línguas. Várias dessas bases têm sido desenvolvidas no formato wordnet, o qual se originou com a construção da base lexical para o inglês norte-americano denominada WordNet de Princeton (WN.Pr) (FELLBAUM, 1998). Em uma base terminológica no formato wordnet (p.ex.: ArchiWordNet, Medical WordNet e Jur-WordNet), as unidades lexicais, ou melhor, as unidades terminológicas (ou termos) de um domínio de especialidade estão organizadas em função dos conceitos a elas subjacentes, os quais são codificados em conjuntos de formas sinônimas (em inglês, synonym set ou synset). Por exemplo: o synset {car, auto, automobile, machine, motorcar} da WN.Pr, construída para a língua geral, codifica o conceito CAR. Os synsets, por sua vez, podem ser relacionados por meio de várias relações conceituais, como hiperonímia/hiponímia, holonímia/meronímia, etc. Para o português do Brasil (PB), no entanto, não há bases de conhecimento especializado no formato wordnet. Nesse cenário, está sendo desenvolvido o projeto TermiNet (do inglês, terminological wordnet), que teve início em 2009 e conta com o apoio financeiro da FAPESP (proc. 2009/06262-1) e do CNPq (proc. 471871/2009-5). Esse projeto objetiva, ao final, construir uma wordnet terminológica (terminet) do domínio da Educação a Distância (EaD) em PB. A essa futura base, dá-se a denominação WordNet.EaD ou WN.EaD. A escolha pela EaD baseou-se especificamente (i) na proeminência desse domínio especializado no atual cenário político e social do Brasil e na (ii) necessidade de uma sistematização vocabular do mesmo. De acordo com a metodologia adotada no projeto TermiNet, a WN.EaD será construída com base em corpus. Sendo assim, todo o conhecimento léxico-conceitual necessário à construção da WN.EaD será extraído semiautomaticamente de um corpus desse domínio, o qual está em plena compilação. Seguindo os pressupostos da Linguística de Corpus, a primeira tarefa realizada no TermiNet foi a projeção do corpus, ou seja, a definição de sua tipologia. Na fase atual de desenvolvimento, os textos que compõem o corpus estão sendo manualmente compilados a partir de material disponível na web. Essa compilação consiste especificamente na identificação de fontes específicas (p.ex.: sites, portais, etc.) e na seleção de textos pertinentes do domínio. Para a seleção das fontes, procedeu-se à delimitação do domínio em questão, ou seja, à identificação dos limites da área-objeto sobre a qual se dá a pesquisa terminológica (BARROS, 2004), posto que, sem conhecer as especificidades da área-objeto não é possível compilar um corpus e, conseqüentemente, identificar a terminologia do domínio e estruturá-la. Neste trabalho, em especial, apresenta-se o que vem a ser a área do Ensino a Distância, não apenas no sentido de apresentar a área-objeto do projeto TermiNet, mas, sobretudo, para demonstrar como se efetua a delimitação de um domínio de especialidade e como essa delimitação guia/ determina a compilação do corpus.