

Extração de informações semânticas de dicionários com vistas à elaboração de léxicos lingüístico-computacionais

Ariani DI FELIPPO
Bento Carlos DIAS-DA-SILVA

Todos os sistemas computacionais que processam língua natural requerem um tipo de “arquivo” em que são armazenadas as unidades lexicais que serão manipuladas pelo sistema durante a interpretação e/ou produção do código lingüístico. Esse “arquivo” é concebido como uma base de dados em que são especificadas, para cada unidade nela contida, informações de natureza lexical, morfológica, sintática, semântica e, até mesmo, pragmático-discursiva. Do ponto de vista computacional, esse tipo de “mega arquivo” é definido como o “léxico” do sistema. Na compilação de léxicos computacionais, os pesquisadores do PLN têm comumente extraído dados léxico-semânticos de obras lexicográficas, posto que estas são inegavelmente fontes de conhecimento lexical. Neste trabalho, em especial, apresenta-se uma breve análise de verbetes extraídos de obras lexicográficas monolíngües do português do Brasil. Mais especificamente, exemplifica-se o processo de compilação, nesses dicionários, de informações léxico-semânticas sobre os substantivos concretos relevantes para a construção de léxicos lingüístico-computacionais.