

"Aquisição Semi-automática de informações lingüísticas: expressões multipalavras e fraseologias"

Daniel Emilio Beck, Aline Villavicencio

Resumo

O Processamento da Linguagem Natural (PLN) é uma área da Ciência da Computação que investiga problemas de geração e compreensão automática de línguas humanas, e sua aplicação é em sistemas computacionais que incluem processamento textual, tais como sistemas de Tradução Automática. A performance destes depende de recursos lingüísticos eletrônicos - como dicionários, tesouros e ontologias - precisos e de alta cobertura. Tais recursos não são completos, dada a magnitude das linguagens naturais e sua natureza dinâmica, com novas palavras e construções aparecendo constantemente, limitando o sucesso de uma aplicação de PLN devido a palavras e especialmente expressões compostas (por exemplo "parada de ônibus) não contidas nestes recursos. Visto que a geração e extensão de recursos lingüísticos manualmente é cara, demorada e sujeita a erros, investigam-se métodos automáticos ou semi-automáticos para a geração desses recursos, a partir de grandes bases de texto (corpora). Desta forma, o objeto de aprendizagem proposto constitui-se de uma ferramenta eletrônica modular e reutilizável onde estão implementadas várias destas técnicas de aquisição automática de conhecimento, proporcionando aos estudantes um ambiente onde podem realizar as tarefas práticas. Além disto, os estudantes tem a possibilidade de contribuir para a continua extensão da ferramenta implementando outras técnicas que potencialmente poderão ser adicionadas a ferramenta.