

SiR/LiCO uma proposta para um sistema de recuperação de Informação baseado em teorias da lingüística computacional e ontologia

GOTTSCHALG-DUQUE, Cláudio. *SiR/LiCO* uma proposta para um sistema de recuperação de Informação baseado em teorias da lingüística computacional e ontologia. 2005. 118f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, UFMG, Belo Horizonte.

Apresenta estudos para a administração de documentos eletrônicos através de um embasamento cognitivista. Propõe uma indexação de textos eletrônicos, disponibilizados em língua portuguesa, por meio da aplicação de teorias de lingüística computacional e utilização de ontologia. A técnica empregada para a indexação é baseada na teoria de Análise proposicional proposta por Frederiksen (1975). Baseia-se na extração de etiquetas sintáticas das palavras que compõem os documentos para a geração de etiquetas semânticas dessas palavras, para gerar uma ontologia leve automaticamente. São sugeridas várias contribuições que visam otimizar o desempenho de sistemas de recuperação de informação, por meio da utilização de técnicas que permitam contextualizar as palavras dos textos a serem indexados. Tais contribuições incluem desde a otimização de analisadores sintáticos, até a geração automática de ontologias leves. Foi criado um *corpus*, uma pequena coleção de documentos eletrônicos sobre ciência da informação, escritos em língua portuguesa e disponibilizados na Web. Esta coleção foi utilizada para testar o protótipo construído. O protótipo, nomeado *SiR/LiCO* (Sistema de recuperação de informação baseado em teorias da lingüística computacional e ontologia), foi utilizado em um experimento-piloto e posteriormente, em outro experimento, com o intuito de verificar e validar a hipótese de que é possível desenvolver e implementar um sistema de recuperação de informação totalmente baseado em teorias lingüísticas, teorias de lingüística computacional e ontologia. Os resultados de precisão e revocação dos experimentos realizados com o *SiR/LiCO* são comparados com os resultados obtidos com a utilização de um modelo vetorial. A análise dos resultados sugere que não só é viável a hipótese defendida como também é muito promissora.