

Ambiente para geração e manutenção semi-automática de tesouros

MOREIRA, Manoel Palhares. *Ambiente para geração e manutenção semi-automática de tesouros*. 2005. 196 f. Tese (Doutorado em Ciência da Informação) — Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte.

Entre as diversas formas de representação da informação utilizadas por Sistemas de Recuperação de Informação encontram-se os tesouros, que se constituem em uma linguagem de indexação consolidada e empregada por profissionais que exercem atividades de organização da informação. A flexibilidade para o estabelecimento de novas relações entre termos, as hierarquias e as referências cruzadas conferem ao instrumento uma multiplicidade de usos, abrangendo processos que vão desde a indexação até a efetiva recuperação dos documentos. A elaboração e manutenção de tesouros são atividades intelectuais com procedimentos específicos, entre eles o conhecimento de documentos produzidos na área, o entendimento dos termos empregados e a construção de conceitos para explicação desses termos. Do profissional envolvido espera-se uma atitude flexível para incorporar as mudanças e inovações que surgem na área, na própria linguagem e no emprego de termos. Este trabalho objetivou a construção de uma metodologia e um ambiente para a geração e manutenção de tesouros de forma semi-automatizada, através da utilização da linguagem natural e com base em tecnologias da Ciência da Computação e nos fundamentos teóricos da Ciência da Informação; mais especificamente, através dos conceitos ordenadores da garantia literária, da garantia de uso e da garantia estrutural, incorporando-se a essas a proposta da garantia advinda da própria estrutura do texto. O ambiente possibilitou a verificação da atualidade e do potencial representativo do tesouro. Partiu-se da hipótese de que palavras-chaves recolhidas de artigos científicos poderiam ser aplicadas neste processo já que elas representam duplamente a garantia literária e a de uso por se tratar de um instrumento privilegiado de disseminação do conhecimento científico. Foram feitos cálculos estatísticos envolvendo frequência e escore padronizado nas observações de frequência de palavras-chave no título, no resumo, no texto e na bibliografia dos artigos. Para testes, foram utilizados textos científicos dos periódicos eletrônicos Datagrama Zero e Ciência da Informação, já consolidados na área. A inexistência de um tesouro atualizado na área levou a construção de um Tesouro em Ciência da Informação (TCI), a partir de tesouros existentes em português (IBICT), em inglês (ASIS) e em espanhol (CINDOC e DOCUTES). O ambiente apontou a necessidade de atualização de termos, classificados por grau de relevância, levando em conta a evolução da área.