

Análise de assunto auxiliada por um sistema de informação, em textos da computação

SILVA, Marilée Patta. *Análise de assunto auxiliada por um sistema de informação, em textos da computação*. 2004. 249f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação da UFMG, Belo Horizonte.

A organização e o armazenamento da informação têm, como objetivo, recuperar informações relevantes aos usuários. Para que um documento seja armazenado, esse deve passar por um processo de indexação, para a sua posterior recuperação. A indexação compreende duas etapas: a análise de assunto, quando conceitos são extraídos do documento e a tradução desses conceitos para uma linguagem de indexação. A análise de assunto é, em geral, uma atividade humana, realizada pelo indexador, profissional inserido em bibliotecas e centros de informação. Atividades como leituras de textos integrais, parágrafos, folhas-de-rosto, autoria, fazem parte da rotina do indexador. Porém, vários fatores influenciam no trabalho. A falta de conhecimento do assunto, dos termos que aparecem no texto, a barreira da língua, a falta de tempo, o volume de trabalho, dificultam tarefa, uma vez que a indexação é uma atividade subjetiva, pois um grupo de indexadores pode atribuir diferentes termos ao mesmo documento. Visando auxiliar a tarefa do indexador, modelou-se e implementou-se um sistema de informação computacional que tem, como objetivo, auxiliar na identificação dos assuntos de textos, colaborando com a análise de assunto. Para tal, fez-se necessária a elaboração de uma base de conhecimento composta por termos de uma área especializada previamente definida, dividida em eixos e sub-eixos. Para cada categoria há a associação de uma lista de termos principais e de termos relacionados, cuja função é auxiliar na identificação dos termos principais. A base de conhecimento contém termos dos eixos de algoritmos e estruturas de dados, de linguagens de programação e de banco de dados, da área da computação. Os textos processados são da mesma área. O sistema de informação processa o conteúdo de textos, comparando-o com os conteúdos dos termos da base de conhecimento. O sistema, ao identificar os termos no texto, calcula a frequência e acumula os pesos e, ao final do processamento, sugere os termos-índices, que refletem os assuntos que predominam no texto por meio do cálculo do percentual de ocorrência do termo no texto, ordenado descendente. Apresentam-se os relatórios de 34 (trinta e quatro) textos. Ao confrontar o resultado do sistema e da análise humana, observou-se que, em todos os textos, foram identificados termos índices coincidentes. Porém, os relatórios fornecem uma lista maior de termos-índices, mapeando, detalhamento, os termos-índices identificados no texto, com os percentuais de ocorrências no texto, ordenados descendente. Caso não haja associação entre os termos da base de conhecimento e do texto, o resultado do processamento retorna vazio.