

Otimização dos processos de indexação dos documentos e de recuperação da informação mediante o uso de instrumentos de controle terminológico

Jaime Robredo

<http://lattes.cnpq.br/9669125022187444>

Publicado em: Ciência da Informação, v. 11 n. 1 jan./jun 1982

Disponível em: <http://revista.ibict.br/ciinf/article/view/175>

RESUMO

Em qualquer área do conhecimento, os termos portadores de significado podem ser utilizados como descritores para representar o conteúdo dos documentos escritos, nos processos de indexação e organização da informação, assim como para formular as perguntas no processo de recuperação da informação. Quando ordenados em função de sua frequência de uso, os descritores apresentam uma distribuição que obedece à lei de Bradford-Zipf. A aplicação das facilidades do processamento eletrônico de dados ajuda grandemente a estabelecer, para áreas específicas do conhecimento, instrumentos de controle terminológico, que permitem otimizar os processos de indexação e recuperação dos documentos, utilizando os termos e as associações entre estes que se destacam por sua riqueza de significado, para representar conceitos determinados. Foram estudados dois universos de termos significativos correspondentes a duas áreas diferentes do conhecimento (agricultura e política científica e tecnológica), resultantes da análise do conteúdo de conjuntos suficientemente grandes de documentos, indexados segundo princípios não subjetivos: indexação automática no primeiro caso e indexação automática simulada no segundo. Desse estudo resultam, para cada caso, listas de descritores estabelecidas a partir de suas respectivas frequências de aparecimento e da aplicação de determinados conceitos de sinonímia e quase-sinonímia, que devem contribuir para otimizar os processos de indexação e recuperação da informação, tanto em sistemas manuais como automatizados. As conclusões estabelecidas no que diz respeito à otimização de indexação parecem confirmadas a partir dos conceitos da teoria da informação.

Palavras chaves: Indexação. Controle terminológico. Recuperação da informação.

Optimization of document indexing and recovery processes through the use of terminological control instruments

ABSTRACT

In any área of knowledge, the meaning-carrier terms may be used as descriptors to represent the content of the written documents in the indexing and Information organizing processes, aswell as to formulate the queries in the information retrieval process. When ranged in function of their frequency of use, the descriptors show a distribution which follows the Bradford-Zipfs law. The data processing facilities may be of great support to prepare, in specif ic áreas of knowledge, Instruments for terminology contrai which enable ones the optimization of the document indexing and retrieval processes by using, to represent specific concepts, the terms and their associations which stand out for their richness in meaning. Two groups of significant terms were studied, corresponding to two different áreas of knowledge (agriculture and scientific and technologic policy), resulting from the contents analysis of sufficiently large sets of documents, indexed according to non-subjective principies: automatic indexing in the first case and simulated automatic indexing in the second. In each case, with basis on their respective occurrences and on the application of synonymy and quasi-synonymy concepts, lists of descriptors were established, which must contribute to optimize the indexing and information retrieval processes, both in manual and automatic systems. The established findings concerning indexing optimization seem confirmed by the information theory concepts.

Keywords: *Indexing. Terminological control. Information retrieval.*

Optimización de índice de documentos y procesos de recuperación a través del uso de instrumentos de control terminológico

RESUMEN

En cualquier área del conocimiento, los términos de portadores de significado se pueden utilizar como descriptores para representar el contenido de los documentos escritos en los procesos de indexación y organización de la información, así como para formular las consultas en el proceso de recuperación de la información. Cuando se clasifican en función de su frecuencia de uso, los descriptores muestran una distribución que sigue la ley de Bradford-Zipfs. Las instalaciones de procesamiento de datos pueden ser de gran ayuda para preparar, en áreas específicas de conocimiento, los instrumentos para la terminología contrai que permiten la optimización de los procesos de indización y recuperación de documentos utilizando, para representar conceptos específicos, los términos y sus asociaciones que se destacan. por su riqueza de significado. Se estudiaron dos grupos de términos significativos, correspondientes a dos áreas diferentes de conocimiento (agricultura y política científica y tecnológica), resultantes del análisis de contenido de conjuntos de documentos suficientemente grandes, indexados según principios no subjetivos: indexación automática en el primer caso. e indexación automática simulada en el segundo. En cada caso, con base en sus respectivas apariciones y en la aplicación de conceptos de sinonimia y cuasi sinonimia, se establecieron listas de descriptores, que deben contribuir a optimizar los procesos de indexación y recuperación de información, tanto en sistemas manuales como automáticos. Los hallazgos establecidos con respecto a la optimización de la indexación parecen confirmados por los conceptos de la teoría de la información.

Palabras clave: *Indización. Control terminológico. Recuperación de información.*