



Proposição, Aplicação e Avaliação de um Método de Classificação Temática em Bases de Dados Textuais Indexadas com Auxílio de Vocabulários Controlados

SERGIO DAGNINO FALCÃO¹

DISSERTAÇÃO DE MESTRADO

Instituição de defesa: *Programa de Pós-Graduação em Gestão do Conhecimento e Tecnologia da Informação da Universidade Católica de Brasília*

Data da defesa: 25/02/2003

PALAVRAS-CHAVE

Gestão do Conhecimento – Sobrecarga de Informações – Descoberta de Conhecimento em Textos – Classificação Temática

RESUMO

A falta de informações foi freqüentemente apontada como fator limitante para a tomada de decisões de forma racional. No entanto, vivemos hoje a Era da Informação e há diversos exemplos de perda de eficiência e de níveis de produtividade abaixo do esperado por parte dos trabalhadores do conhecimento, devido, em parte, ao excesso de informações a que são expostos cotidianamente.

Este trabalho descreve a proposição, a aplicação e a avaliação de um método de classificação temática em uma base de dados com discursos proferidos por deputados federais no Plenário da Câmara dos Deputados da República Federativa do Brasil entre outubro de 2000 e outubro de 2002 e que foi indexada com auxílio de um vocabulário controlado. O método desenvolvido utiliza os recursos de um banco de dados relacional para atribuir temas aos discursos, por meio da análise dos descritores utilizados na indexação. Os 10.627 discursos foram agrupados em 14 temas; foram feitas subdivisões da classificação temática por região geográfica e por partido político do orador, também ao longo do tempo e a possível correlação entre os temas.

A aplicação do método de classificação temática foi avaliada por 36 funcionários da Câmara dos Deputados, envolvidos profissionalmente com o assunto da base dados analisada, os quais foram entrevistados por meio de questionário. Verificou-se que a aplicação do método proposto permite contextualizar as informações armazenadas, agregando-lhes valor através da atribuição de significado e propósito. Constatou-se ainda que a aplicação do método

¹ E-mail: sfalcao@terra.com.br

possibilita a descoberta de conhecimentos através da identificação de padrões válidos, novos e potencialmente úteis nas informações armazenadas. Foram também relacionadas possíveis utilizações da aplicação do método na base de dados completa ou em outras bases de dados.

KEYWORDS

Knowledge Management - Information Overload - Knowledge Discovery in Texts - Subject Classification

ABSTRACT

The lack of information has been often appointed as limiting factor with respect to the decision making process in rational ways. However, we live today in the Information Age and there are many examples of loss of efficiency and levels of productivity below expected by the so called knowledge workers, due to, in part, the information overload they are exposed to. This work describes the proposal, deployment and evaluation of a thematic classification method in a full text database containing speeches pronounced by representatives in the plenary assembly of the Chamber of Deputies of the Federative Republic of Brazil, between October 2000 and October 2002, and that was indexed with aid of a controlled vocabulary. The method developed uses the features of a relational data base to assign subjects to the speeches, through the analysis of the terms used in the indexation field. The 10,627 speeches had been grouped into 14 main subject categories and it had been made subdivisions of thematic classification by geographic region and political party of the speaker, and also by the date of the speech as well as the possible correlation between subjects. The deployment of the method of thematic classification was evaluated by 36 employees of the Chamber of Deputies, professionally involved with the subject of the database being analyzed, who had been interviewed by means of questionnaires. It was verified that the deployment of the proposed method allows the contextualization of the stored information, adding value to it through the assignment of meaning and intention. It was also evidenced that the deployment of the proposed method makes possible the process of knowledge discovery through identification of valid, new and potentially useful patterns in the stored data. Possible uses of the deployment of the method in the complete database or other databases had been related.

SOBRE O AUTOR

Sérgio DagninoFalcão

Bacharel em Ciência da Computação pela Universidade de Brasília (UnB) e Mestre em Gestão do Conhecimento e Tecnologia da Informação pela Universidade Católica de Brasília (UCB). É membro da Câmara Técnica de Documentos Eletrônicos do Conselho Nacional de Arquivos – CTDE/Conarq. Participa como pesquisador do Projeto InterPARES 3 – International Research on Permanent Authentic Records in Electronic System. Atualmente trabalha na Assessoria do Centro de Informática da Câmara dos Deputados – O texto da dissertação encontra-se disponível em http://www.bdtb.uceb.br/tede/tde_busca/arquivo.php?codArquivo=161