

SILVA, Edson Marchetti da. Recuperação da informação através de busca comparada em domínio específico, baseado em expressões multipalavras. 177f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

*Normalmente, as ferramentas de busca em bases de dados utilizam-se de palavras-chave fornecidas pelo usuário para realizar a identificação de documentos. Este trabalho visa propor uma alternativa adicional que possa ser agregada aos Sistemas de Recuperação da Informação (SRI) para auxiliar o usuário em seu processo de busca pela informação em base de documentos. Essa alternativa possibilita a realização de uma busca automatizada baseada em um documento fornecido pelo usuário que serve de referência. Nesse contexto, delimitou-se como objeto de estudo a extração de expressões multipalavras (EM) do documento para servir como descritores da busca em um corpus específico. As EM são obtidas através de um método determinístico proposto que considera as características da estrutura física do documento e compara o resultado com o obtido por treze diferentes medidas de associação estatísticas produzidas pelo software Ngram Statistics Package (NSP) que considera o texto como um conjunto de palavras do inglês bag of words. Os resultados demonstram que o método proposto possibilita uma melhor representação semântica do documento trazendo ganhos qualitativos no conjunto de EM extraídas e que contribui positivamente para o resultado da Busca Comparada. A partir desses experimentos propôs-se e implementou-se um protótipo de ferramenta de Busca Comparada e apresentam-se os resultados obtidos com o seu uso.*