

Recuperação de Informação e Expansão Automática de Consulta com Thesaurus: uma avaliação

Marco Gonzalez

PUCRS - Faculdade de Informática
Av.Ipiranga, 6681 – Prédio 16, térreo
90619-900 Porto Alegre, Brazil
gonzalez@inf.pucrs.br

Vera Lúcia Strube de Lima

PUCRS - Faculdade de Informática
Av.Ipiranga, 6681 – Prédio 16, térreo
90619-900 Porto Alegre, Brazil
vera@inf.pucrs.br

ABSTRACT

This paper presents a first evaluation of obtained results on automatic query expansion in information retrieval (IR). A thesaurus with semantic structuring and generative operations is used for deriving the lexical set that is related to each query term for automatic expansion. The selection of new terms of the expanded query and the calculation of its weights depend on the intersection of the derived lexical sets and on the depth level for descriptors search with respect to each considered terms. In this evaluation the following issues were considered: (i) a dissertations summaries collection about information systems; (ii) index terms in canonical form; and (iii) a group of test queries where the terms are key-words associated (by a traditional library system) with the dissertations. The results indicate that the query expansion, with the used approach, brings benefits for IR.

Keywords: Information Retrieval, Natural Language Processing, Artificial Intelligence, Thesaurus, query expansion, Data Mining

RESUMO

Este artigo apresenta uma primeira avaliação dos resultados obtidos com a expansão automática de consulta em recuperação de informação (RI). Um thesaurus, com estruturação semântica e operações gerativas, é utilizado para gerar o campo lexical de cada termo da consulta e obter a expansão automaticamente. A seleção dos novos termos e o cálculo de seus pesos, na consulta expandida, depende da sobreposição dos campos lexicais e do nível de profundidade que se avança na busca de descritores dos termos considerados. Foram utilizados nesta avaliação: (i) uma coleção de documentos constituída por resumos de dissertações da área de sistemas de informação; (ii) termos de índice em forma canônica; e (iii) um conjunto de consultas de teste oriundas de palavras-chave associadas (por um sistema tradicional de biblioteca) aos documentos. Os resultados indicam que a expansão de consulta, com a abordagem utilizada, traz benefícios para a RI.

Palavras-chave: Recuperação de Informação, Processamento de Linguagem Natural, Inteligência Artificial, Thesaurus, Expansão de Consulta, Data Mining