

Resumo

Com o massivo aumento da disponibilização de novos conteúdos na Internet, a pesquisa de informação tomou-se cada vez mais importante. Desde o seu início que os sistemas têm vindo a sofrer constantes desenvolvimentos e a ser alvo de investigação por parte de uma vasta comunidade científica.

Derivado do nosso estudo, implementamos um novo sistema de agrupamento e apresentação de resultados que advém da procura da informação em ambiente web. Utilizamos técnicas de Web Content Mining para representar os textos a partir dos seus termos mais relevantes, que tanto podem ser palavras simples como palavras compostas. No seguimento, aplicamos um algoritmo de soft clustering que agrupa em clusters documentos relativos ao mesmo conceito, apresentando-os de forma hierárquica, labelizados com os seus termos mais relevantes. Este sistema evoluirá para uma nova vertente de Query Expansion, denominada de Classified Query Expansion. Será Automática (acrescentando termos à query) com a implementação progressiva de uma WebWarehouse (que guardará os termos relacionados a cada nova pesquisa efectuada) e será Interactiva (sugerindo ao utilizador um conjunto de resultados que se apresentam de forma hierárquica), propondo ao utilizador a expansão da query escolhendo um ou mais clusters.

A organização dos resultados, que decorre deste novo conceito facilitará a navegação do utilizador pela lista de páginas devolvidas por um motor de busca qualquer. Com uma ferramenta que automatize estes processos, o utilizador não mais necessitará de fazer uma procura exaustiva da página do seu interesse, o que significa um considerável ganho de tempo que o mesmo poderá dedicar a outras tarefas como sejam o estudo em concreto do site do seu interesse.