



XVI congresso interno de iniciação científica

Ginásio Multidisciplinar da Unicamp
24 a 25 de setembro de 2008



E0384

RECUPERAÇÃO DE DOCUMENTOS UTILIZANDO RELEVANCE FEEDBACK BASEADO EM PROGRAMAÇÃO GENÉTICA

Enrico Silva Miranda (Bolsista PIBIC/CNPq) e Prof. Dr. Ricardo da Silva Torres (Orientador), Instituto de Computação - IC, UNICAMP

Grandes repositórios de documentos textuais (teses, dissertações, jornais, etc.) vêm sendo criados recentemente. Este cenário faz com que haja uma demanda por sistemas que permitam uma busca eficiente de documentos relevantes. Os sistemas comuns de busca por palavras-chaves podem retornar resultados irrelevantes do ponto de vista do interesse de usuário. Técnicas de realimentação de relevância tentam contornar este problema. Basicamente estas técnicas utilizam a interação do usuário (indicação de documentos relevantes) com resultados de uma consulta para “aprender” quais documentos são relevantes. Com isso, o sistema de busca se adapta melhor às necessidades diferenciadas de cada usuário. Esta iniciação científica objetiva desenvolver uma forma de recuperar estes documentos usando relevance feedback baseado em programação genética. Programação genética é uma técnica de aprendizado baseada em conceitos biológicos de seleção natural. Ela utiliza conceitos como mutação, crossover e reprodução genética para que uma população de indivíduos (soluções do problema-alvo) se adapte às condições dadas (neste caso, indicação de documentos relevantes). O estágio atual da pesquisa compreende a implementação de um protótipo e testes iniciais em repositórios pequenos contendo teses e dissertações da Unicamp. O protótipo encontra-se em estágio avançado de implementação com o módulo de similaridades já desenvolvido e o módulo de Relevance Feedback baseado em Programação Genética em fase final de desenvolvimento. Finalizada esta última etapa, pretende-se incorporar o mecanismo de relevance feedback ao sistema Nou-Rau usado na Biblioteca Digital de Teses e Dissertações da UNICAMP.

Recuperação de documentos - Relevance feedback - Programação genética