



UNIVERSIDADE FEDERAL
DE MINAS GERAIS

PROGRAMAÇÃO E DESENVOLVIMENTO DE SOFTWARE II - (PDSII)

Trabalho Prático: Máquina de Busca

Vitor Marques de Souza

Introdução:

Uma máquina de busca é nome dado aos robôs que são capazes de retornar uma lista de documentos mais relevantes de forma ordenada de acordo com uma pesquisa específica. Existem diversos passos a serem seguidos para sua implementação e neste relatório mostra-se um pouco do que foi feito no projeto.

Implementação:

O primeiro passo foi a criação da representação dos documentos que corresponde a criação do índice invertido, o modelo de estrutura de dados adotado foi o de um map em que cada palavra contida nos documentos foi usada como chave ("Key") e para os valores foi usado um set com os documentos nos quais as palavras aparecem. A escolha dessa estrutura foi a mais adequada pois precisávamos guardar 2 tipos de informações: as palavras e os documentos; todos os outros tipos de estruturas sozinhos como vector e list não seriam adequados e eficientes para a associação de 2 tipos diferentes de informação.

Em seguida discutiu-se a forma como iria ser lida as palavras contidas no documento. Para que não houvesse repetição da palavra dentro do índice, retirou-se todas as palavras como strings (devidamente convertidas) e as armazenou dentro de um set (pois nesse tipo de estrutura não há repetição), logo após foram inseridas dentro do map e para armazenar os documentos usou-se um outro laço que, como o valor do map é a classe set, a propriedade de inserir o nome do documento foi a sintaxe mais adequada.

A classe índice possui funções escolhidas para que se possa retirar as informações necessárias contida nela, como número de documentos, tamanho do índice e etc.

Em seguida temos a criação da classe ranking cosseno. Para armazenar as coordenadas de cada documento criou-se uma outra estrutura de dados, foi escolhido um map para guardar o nome de cada documento como chave e um vector para guardar as coordenadas de cada documento como valor. Desse modo a função de similaridade juntamente com a função rankear calcula e armazena em outra estrutura de dados os documentos mais próximos daquele buscado pelo usuário do programa. Todas as outras funções são suportes e ou formulas intermediarias para se criar o ranking cosseno, foram todas criadas como instruído no pdf do trabalho.

Conclusão:

Com esse trabalho conclui-se a importância de se fazer um código bem modularizado e como isso ajuda na sua manutenção e na produtividade dos programadores. Ademais, pudemos utilizar todas as técnicas de programação mais avançada no projeto tais como testes de software, containers de dados e a própria orientação a objeto.