UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMAÇÃO E DESENVOLVIMENTO DE SOFTWARE II

Trabalho Prático

Gabriel Teixeira Lacerda
2016026302
Richard Wagner Abras Ribeiro Sobrinho
2013030244

A máquina de busca

O problema da máquina de busca proposto neste trabalho propõe a ação de se ler arquivos de texto para buscar neles uma palavra. Para isso, foi preciso construir uma associação entre ela e os arquivos em que foi encontrada, listando-os na sequência.
A apresentação do resultado da busca é em forma de índice invertido. Ao contrário de um índice comum, aqui cada entrada de busca aparece vinculada à lista ordenada dos documentos que a contêm.

Desenvolvimento

Parte 1

Leitura de arquivos
O usuário deve digitar o nome de cada arquivo para se estabelecer o conjunto de entrada. Na sequência, é criado um vetor contendo os nomes dos arquivos, que serão adicionados em loop até que o usuário digite “fim”. Uma verificação foi adicionada para que os arquivos inexistentes, ou ainda com nomes digitados incorretamente, sejam apontados.
Todas as palavras dos documentos são lidas para verificação, passando por um filtro, mantendo apenas os caracteres alfanuméricos e convertendo as letras maiúsculas em minúsculas, utilizando a função implementada “removePunctuation” com base na verificação “iswalnum” e a função “tolower”.
Somente após essas alterações é criado, de fato, o índice invertido.

Parte 2

Estrutura de dados do índice invertido
Seguindo seu próprio conceito, foi preciso criar um mapa de chaves (todas as palavras contidas nos documentos) com seus valores associados (cada conjunto com os nomes dos documentos que contêm cada palavra).
As palavras nos documentos são, uma a uma, filtradas conforme as condições supracitadas e logo em seguida inseridas no índice invertido, formando o mapa com os seus valores (set). O

nome do arquivo da palavra é, então, inserido no set de strings correspondente de tal forma que não seja adicionado em redundância.

Parte 3

Consultas
A consulta de apenas uma palavra permite uma verificação simples e direta nos valores de chave do índice invertido. Com a correspondência aos sets, os resultados são impressos em ordem alfabética. Caso a palavra não seja encontrada, o usuário recebe uma mensagem.

Conclusão

A execução do trabalho se deu de maneira simples, quando pudemos sanar algumas dúvidas práticas. A teoria por trás da construção do índice invertido não foi uma barreira, ao contrário do próprio uso do C++. Também esbarramos em dificuldades com a compreensão dos testes de unidade e, neste caso, creio que o ponto inicialmente positivo de encontrar diversos frameworks disponíveis também afetou negativamente a minha segurança para usá-los. Além disso, a forma como começamos a construir o código dificultou a separação em unidades para os testes, conforme nosso entendimento. Tentamos compensar essa questão com verificações intermediárias, dando mais segurança no desenvolvimento e dando retornos claros, diretos e imediatos às entradas.