

**Rafael Mendes de Souza2014015559**  
**Gabriel Breunig de Souza Nunes 2104430521**

## 1. Introdução

Um buscador é um programa projetado para procurar palavras fornecidas pelo usuário em documentos e bancos de dados. No contexto da internet, ele permite procurar palavras-chave em documentos, como aqueles que se encontram armazenados em sites.

O objetivo deste trabalho é criar um programa com princípio de funcionamento semelhante ao dos buscadores para encontrar palavras digitadas pelo usuário em um conjunto de arquivos no formato (txt).

## 2. Implementação

O princípio de funcionamento de buscadores, tais como Google e Yahoo!, é baseado na reunião de documentos para construção de um índice invertido. O índice invertido usa o princípio da associação de tal forma que cada palavra é associada a um conjunto de documentos onde ela está presente.

### Estrutura de Dados

Para a implementação do índice invertido foi usada uma estrutura tipo mapa, uma estrutura associativa na qual os elementos armazenados estão organizados na forma de pares (chave, valor), de tal forma que é possível ter acesso a valor (que pode eventualmente ser um objeto com uma estrutura complexa) obter o valor a partir da especificação da chave.

Em C++, a biblioteca STL implementa mapas com a classe **map**. A declaração de uma coleção deste tipo especifica dois parâmetros, o primeiro para o tipo da chave e o segundo para o tipo do valor.

A estrutura **map** não admite duplicação de chaves; STL oferece também uma implementação **multimap**, na qual as chaves podem ser repetidas, essa estrutura foi usada porque a palavra procurada pode existir em mais de um arquivo, sendo que a palavra procurada é a chave e os arquivos onde ela está são os valores.

Outra estrutura utilizada foi um vetor dinâmico da biblioteca STL **vector**, esse é uma coleção de dados de um mesmo tipo que aumenta de tamanho de forma dinâmica de acordo com a necessidade do programa em execução. Ele foi utilizado para armazenar os valores (nomes dos arquivos) associados as chaves no momento da impressão na tela.

### Funções e Procedimentos:

A seguinte função foi usada (**multimap**):

**insert(pair <string, string> (chave, valor):** Insere uma chave e um valor na estrutura onde ambas são strings.

A seguinte função foi usada (**vector**):

**push\_back(valor):** Insere um valor no vetor dinâmico onde esse valor é uma string.

### Programa Principal

O programa principal cria uma variável do tipo **buscador**, classe onde contem os métodos para operar o índice invertido, em seguida recebe uma palavra enviada pelo usuário. Então o programa faz a busca no conjunto de arquivos pré definidos guarda no índice invertido busca no mesmo guarda no vector então imprime na tela.

## **Organização do Código, Decisões de Implementação e Detalhes Técnicos**

O código está dividido em quatro arquivos principais: *mian.cpp*, *buscador.cpp*, *buscador.h* e *util.h*. implementam o Tipo

### **4. Teste**

Vários testes foram realizados com o programa de forma a verificar o seu funcionamento. Os testes foram realizados em um Pentium Dual-Core, com 2 Gb de memória.

Testes como, palavras com caracteres especiais e letras maiúsculas e minúsculas de forma desordenada, não causam nem um problema na execução do programa.

### **5. Conclusão**

A implementação do trabalho transcorreu sem maiores problemas e os resultados ficaram dentro do esperado.

A principal dificuldade encontrada foi pensar em uma organização para as funções e métodos de forma a criar uma unidade não um conjunto de arquivos separados.

## **Referências**

[1] - [https://pt.wikipedia.org/wiki/Motor\\_de\\_busca](https://pt.wikipedia.org/wiki/Motor_de_busca)

[2] - <http://lampiao.ic.unicamp.br/maratona/?name=estruturas>