

Aula 13: Trabalho 3

Prof.: Paulo Roberto Nunes de Souza

15 Trabalho 3 - PATRICIA

Desenvolver um programa que implementa uma árvore PATRICIA para efetuar a busca de strings. O programa receberá uma lista palavras a serem buscadas em todos os arquivos ".txt" do diretório atual.

16 Execução do programa

O programa deve receber o nome do arquivo de entrada na linha de comando, conforme o seguinte formato:

```
<programa> <arquivo_com_palavras_buscadadas>
```

Supondo que o nome do arquivo executável criado seja "patricia" e o nome do arquivo com a lista de palavras buscadas seja "procuradas.dat", a linha de comando ficaria da seguinte forma:

Execução do programa

```
1 ./ patricia procuradas.dat
```

16.1 Arquivo de entrada: lista de palavras a serem buscadas

A lista das palavras a serem buscadas será passada por arquivo e para cada palavra do arquivo deve-se efetuar a busca na árvore.

Em cada linha do arquivo existirá apenas uma palavra para ser buscada. O arquivo seguirá o seguinte formato:

```
palavra_1
palavra_2
palavra_3
palavra_4
...
palavra_n
```

Veja a seguir o exemplo de um arquivo de entrada com lista de palavras a serem buscadas.

Lista de palavras a serem buscadas

```
4 test
new
news
dog
```

Para fins de simplificação, considere que só serão procuradas palavras apenas com letras sem acentuação, sem caracteres especiais e sem números.

17 Indexação

O programa ao iniciar a sua execução, deve percorrer o diretório corrente[1] identificando todos os arquivos com extensão ".txt". Cada arquivo ".txt" encontrado deve então ser aberto e percorrido. Ao percorrer cada arquivo o programa deverá inserir cada palavra encontrada na árvore PATRICIA.

Para fins de simplificação, considere que só serão consideradas como palavras no texto, as substrings formadas apenas por letras sem acentuação, sem caracteres especiais e sem números.

18 Resultados

Para cada palavra buscada o programa deve retornar a palavra que está sendo buscada, **sem diferenciar letras minúsculas de maiúsculas**, e todas as ocorrências da palavra em todos os arquivos ".txt" do diretório, caso exista alguma, caso contrário deve aparecer a expressão "Palavra não encontrada". Para identificar cada ocorrência encontrada o programa deverá retornar as seguintes informações:

- arquivo onde a ocorrência foi encontrada;
- trecho de texto com a ocorrência, iniciando-se 10 caracteres antes e terminando 20 caracteres após a primeira letra da palavra buscada;

Veja a seguir o exemplo da saída esperada do programa.

Saída do programa de busca

```

1 test
  Palavra não encontrada
new
  texto3.txt: to use a new design two weeks
  texto3.txt: that the new design is better
6  texto3.txt: thank our new design company.
  texto3.txt: and reach new clients through
  texto3.txt: ttracting new customers. Webno
  texto4.txt: choose a new speaker of the p
  texto5.txt: ded for a new home at a primat
11 news
  texto3.txt: ou with a news website. This i
dog
  Palavra não encontrada

```

O programa deve ser escrito na linguagem de programação Ansi-C. Os programas submetidos serão testados usando-se o compilador GCC no sistema operacional Linux.

19 Critérios de correção

O trabalho será corrigido segundo 5 critérios:

P Ausência de plágio

A Adequação ao solicitado

C Correção dos resultados

E Entrega no prazo

D Desempenho computacional

Para a obtenção da nota do trabalho, cada critério receberá um valor entre 0 e 1 seguindo os testes efetuados pelo professor. A nota será o resultado da seguinte ponderação:

$$Nota = P \times A \times (7 \times C + E \times (2 + D))$$

20 Grupo

Os trabalhos devem ser feitos em grupos de até 2 pessoas.

21 Entrega

Os resultados devem ser entregue em meio digital até o dia **23/07/2013** pelo e-mail **paulonunes@cca.ufes.br**.

Referências

- [1] Free Software Foundation, Inc. The GNU C Library. http://www.gnu.org/savannah-checkouts/gnu/libc/manual/html_node/Simple-Directory-Lister.html#Simple-Directory-Lister