

Universidade Federal do Amazonas Faculdade de Tecnologia

Trabalho 3: Árvore PATRICIA

Manaus

Árvore Patrícia

Trabalho prático apresentado à disciplina Algoritmos e Estruturas de Dados como parte dos requisitos necessários para obtenção de nota parcial.

Professor: Edson Nascimento Silva Júnior

Manaus

INTRODUÇÃO

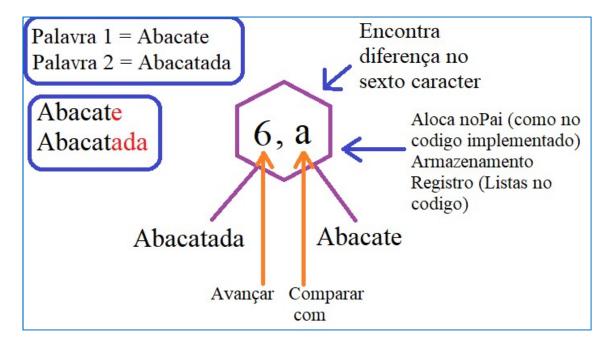
Nas Arvores Trie, definidas em 1960 por Edward Fredkin, termo advindo do inglês, retrieval, cada nó contém informações sobre um ou mais símbolos do alfabeto utilizado. Elas são estruturas indicadas para suportar tarefas de tratamento léxico, como: buscas em dicionários, pesquisas em textos grandes, índices de documentos, entre outros padrões de pesquisa.

A utilização de uma Trie apenas é eficiente se o acesso aos componentes individuais das chaves for bastante rápido. Quanto maior a estrutura mais eficiente uso do espaço. Para driblar o desperdício de espaço com estruturas pequenas foram criadas as árvores PATRICIA

Definida em 1968 por Donald R. Morrison, a Trie Compactada Binária, ou PATRICIA (acrônimo de Pratical Algorithm To Retrieve Information Coded in Alphanumeric), é particularmente útil para tratar de chaves de tamanho variável extremamente longas, tais como títulos e frases.

Também conhecida como RADIX, É uma representação compacta de uma Trie, onde os nós que teriam apenas um filho são agrupados nos seus antecessores. É comum que muitos nós das tries tenham apenas um nó descendente. Isto faz com que as tries tenham um custo elevado de espaço Caminhos que possuem nós com apenas 1 filho são agrupados em uma única aresta. Diferente das Tries não armazena informações nos nodos internos, apenas contadores e ponteiros para cada subárvore descendente.

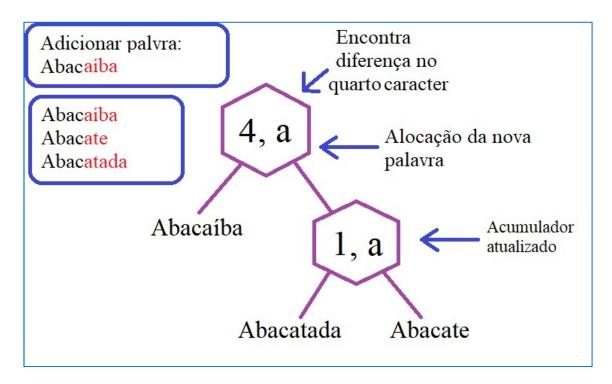
A seguir, esquematização para demonstrar brevemente a dinâmica da PATRICIA:



Exemplo pela alocação das palavras "abacate" e "abacatada"

O campo "avançar" é um registro acumulativo que integra todos os nós, exceto os folhas. Identifica qual a posição do caracter da chave informada que deve ser analisado. O campo "comparar com" apresenta o caracter que deve ser comparado ao caracter da

chave informada como nas árvores binárias de busca, após a análise, se a chave é menor ou igual ao nodo ela é alocada na esquerda ou na direita.



Exemplo pela inserção de nova palavra

Primeiro nó informando pra comparar 4º caracter da palavra. Logo depois, como "t" é maior que "i" desloca-se pra sub-árvore da direita. Então comparam-se agora as ultimas duas palavras.

Mais detalhes no tópico da Implementação do código.

IMPLEMENTAÇÃO

1. Enunciado

O programa foi criado para ler um texto qualquer (arquivo no formato .txt) e imprimir, em ordem alfabética, todas palavras com 3 ou mais caracteres e a(s)linha(s) na(s) qual(is) elas aparecem no texto, desprezando expaços em branco e sinais de pontuação, considerando-os separadores de palavras. Além disso, a leitura deverá converter todas as letras maiúsculas em minúsculas, considerando cada que palavra contém no máximo 20 letras e que não haverá caracteres acentuados.

Exemplo de entrada:

Um		
Dois		
Teste		
Testando		
Quatro		
Ultimo teste		
Adeus		

Saída:

Adeus, linha 7
Dois, linha 2
Quatro, linha 5
Testando, linha 4
Teste, linha 3, linha 6
Ultimo, linha 6

2. Programa

a. Estrutura

Foram declaradas as estruturas e os protótipos das funções em uma biblioteca.h (arvP.h), funções implementadas no arquivo .c (arvP.c) e a main com leitura dos arquivos e manipulação de caracteres em arquivo .c (mainP.c).

Vale ressaltar que no código, foram manipuladas inúmeras estruturas adquiridas como conhecimento ao longo do curso, nas matérias de Algoritmos e Estruturas de Dados (tanto I quanto II), tais como árvore, vetores, listas, ponteiros, condicionais, entre outros, além de serem incluídas diversas bibliotecas, permitindo que o trabalho fosse um ótimo meio para revisão, consolidação e finalização dos aprendizados desta etapa.

```
//declaraçao de um tipo String
typedef char* String;
//passando como parâmetro a enumeração dos tipos de NO para ser retornada em funcao futuramente
 Pai, Folha //noPai (interno) possui filhos, noFolha (externo) sem filhos
//estruturas para Lista, usada na implementacao
typedef struct elemento{
   int dado;
   struct elemento *proximo;
}tipoEle;
typedef struct lista{
   tipoEle *primeiro;
tipoEle *ultimo;
}tipoLista;
//estruturas da arvore e seus nos
typedef struct tNoPAT *tipoPAT;
typedef struct tNoPAT{
  tipoNo tipo;
      larar dentro de uma uniao: deixando todos os campos na mesma posição de memória
       struct{
         short Indice;
char Caractere;
         tipoPAT Esquerdo, Direito;
      }noPai;
      String Dado;
   }NO;
   tipoLista *linhas; //Guarda as linhas que aparece Chave
}tNoPAT;
//prototipos das funcoes
char Digito(short, String);
int Comparou(char, char);
int EFolha(tipoPAT);
void Ordena(tipoPAT);
void ignoraSimbolo(String, int);
void inserirLista(tipoLista *, int);
void imprimirLista(tipoEle *);
void mostrarPalavra(tipoPAT);
tipoPAT criaPai(short, tipoPAT *, tipoPAT *, char, int);
tipoPAT criaFolha(String, int);
tipoPAT Busca(String, tipoPAT);
tipoPAT InsereAux(String, tipoPAT, short, int);
tipoPAT Insere(String, tipoPAT, int);
tipoLista *criarLista();
```

b. Inserção

Foi usada uma função auxiliar de alocação para cada nodo e função principal para atribuir valores, e a função comparação essencial para o código.

```
tipoPAT InsereAux(String palavra, tipoPAT arv, short i, int linha){
  tipoPAT no;
  char caractere;
 if (eFolha(arv)
                    || i < arv->NO.noPai.Indice){
    if(eFolha(arv)){
      //se identificou folha, retorna o i caractere = Digito(i,arv->NO.Dado);
    }else{
       caractere = arv->NO.noPai.Caractere;
    no = criaFolha(palavra, linha);
    if (Comparou(Digito(i, palavra), caractere)){
       return (criaPai(i,&no, &arv,Digito(i, palavra), linha));
      return (criaPai(i,&arv,&no,caractere, linha));
  }else{
    if(Comparou(Digito((arv)->NO.noPai.Indice, palavra),(arv)->NO.noPai.Caractere)){
  (arv)->NO.noPai.Esquerdo = InsereAux(palavra,(arv)->NO.noPai.Esquerdo,i, linha);
      (arv)->NO.noPai.Direito = InsereAux(palavra,(arv)->NO.noPai.Direito,i, linha);
    return (arv);
```

```
tipoPAT Insere(String palavra, tipoPAT arv, int linha){
  tipoPAT no;
  int i;
  if(arv == NULL){
    return (criaFolha(palavra, linha));
   no = arv;//faz o no ser raiz da arv
   while (!eFolha(no)){
     if(Comparou(Digito(no->NO.noPai.Indice,palavra),no->NO.noPai.Caractere)){
       no = no->NO.noPai.Esquerdo;
      }else{
       no = no->NO.noPai.Direito;
   }i = 0;
    if(strlen(palavra)>strlen(no->NO.Dado)){
      while ((i <= strlen(palavra)) && (Digito(i, palavra) == Digito(i, no->NO.Dado)))
       i++;
if (i >= strlen(palavra)){
  inserirLista(no->linhas, linha);//inserir valorLinha na linha e retornar raiz
        }else{ //inserir palavra
          return (InsereAux(palavra, arv, i, linha));
    }else{while((i <= strlen(no->NO.Dado)) && (Digito(i, palavra) == Digito(i, no->NO.Dado)))
       i++;
if (i >= strlen(no->NO.Dado)){
          inserirLista(no->linhas, linha);
        return (arv);
}else{return (InsereAux(palavra, arv, i, linha));
```

c. Busca

Para a busca necessária para identificar as chaves e armazenar suas linhas, foi utilizada uma lista para registro.

```
//busca palavra e verifica se existe na arvore
tipoPAT Busca(String palavra, tipoPAT arv){
   if(eFolha(arv)){
      if(strncmp(palavra, arv->NO.Dado, strlen(palavra)) == 0){
        return arv;
   }else{
      return NULL;
   }
   }
   if (Comparou(Digito(arv->NO.noPai.Indice, palavra),arv->NO.noPai.Caractere)){
      return Busca(palavra, arv->NO.noPai.Esquerdo);
   }else{
      return Busca(palavra, arv->NO.noPai.Direito);
   }
}
```

```
tipoLista *criarLista(){
   tipoLista *aux = malloc(sizeof(tipoLista));
   aux->primeiro = NULL;
   aux->ultimo = NULL;
   return aux;
//insere valores na lista
void inserirLista(tipoLista *lista, int dado){
 if(lista->primeiro == NULL){
   tipoEle *aux = malloc(sizeof(tipoLista));
   aux->proximo = NULL;
    aux->dado = dado;
    lista->primeiro = aux;
  }else{
    tipoEle *temp = lista->primeiro;
   while(temp->proximo != NULL){
     temp = temp->proximo;
   tipoEle *aux = malloc(sizeof(tipoEle));
   aux->dado = dado;
   aux->proximo = NULL;
   temp->proximo = aux;
    lista->ultimo = aux;
  }
void imprimirLista(tipoEle *linha){
 if(linha == NULL){
   return;
 }else{
   printf(" - na linha %d;", linha->dado);
    imprimirLista(linha->proximo);
  }
```

d. Função principal

Contém as linhas de comando para fazer a chamada e a leitura de um arquivo texto com manipulação adequada das strings.

```
sinctude cxtdio.hn
strip (lefe cxtdio.hn)
strip frace a salloc(sizeof(char)*21);
string frace a salloc(sizeof(char)*21);
string letras = strtok(frase, "");
while(letras = strtok(frase, "");
strtok(caractere, letras);
//precurse u exos caractere;
//pr
```

CONCLUSÃO

Como já dito, em geral, há grande importância neste trabalho em relação ao entendimento de toda uma etapa referente aos estudos das disciplinas Algoritmos e estruturas de Dados.

A princípio, entender a forma de execução da PATRICIA foi um desafio. A procura por visualizações ou um norte para implementação foi difícil, mas com a ajuda do professor, numa aula emergencial de dúvidas, muito foi esclarecido, e as dicas ajudaram bastante.

Existiram dificuldades nas questões da manipulação dos arquivos e das strings, como ordenação de caracteres e remoção de caracteres indesejados.

Também na parte da contagem das linhas, foi difícil pensar e concluir uma ideia decisiva para a contagem dentro da estrutura da árvore PATRICIA. Desafios à parte, a implementação foi concluída e o programa executado como desejado.

Chamada: > prog texto.txt | . Segue execução do programa:

Primeiramente, fiz um texto teste para demonstração do modelo da execução;



```
icrosoft Windows [versão 10.0.19041.572]
:\Users\Laura>cd desktop\arvore2
::\Users\Laura\Desktop\arvore2>gcc mainP.c arvP.c arvP.h -o programaPATRICIA
:\Users\Laura\Desktop\arvore2>programaPATRICIA textoTeste1.txt
kamila - na linha 27;
:\Users\Laura\Desktop\arvore2>
```

Agora, execução com os textos indicados no trabalho:

a. Faroeste.txt

```
:\Users\Laura\Desktop\arvore2>gcc mainP.c arvP.c arvP.h -o programaPATRICIA
:\Users\Laura\Desktop\arvore2>programaPATRICIA faroeste.txt
acabou - na linha 66;
ainda - na linha 12;
ali - na linha 150; - na linha 152;
alta - na linha 88; - na linha 169;
apareceu - na linha 112;
aprendiz - na linha 45;
aquilo - na linha 18;
```

```
arma - na linha 188;

arma - na linha 185;

arma - na linha 186;

arma - na linha 187;

bandari - na linha 187;

arma - na linha 187;

bandari - na linha 187;

bangari - na linha 187;

bangaria - na linha 187;

carriari - na linha 187;
```

b. Brsail.txt

```
:\Users\Laura\Desktop\arvore2>gcc mainP.c arvP.c arvP.h -o programaPATRICIA
:\Users\Laura\Desktop\arvore2>programaPATRICIA brasil.txt
armaram - na linha 6; - na linha 30;
assim - na linha 23; - na linha 47; - na linha 60; - na linha 68;
cartao - na linha 18;
cazuza - na linha 2;
chefe - na linha 17;
cigarro - na linha 13;
convencer - na linha 31;
convidaram - na linha 28;
confia - na linha 51; - na linha 64; - na linha 72; - na linha 73;
essa - na linha 9; - na linha 33;
esta - na linha 5;
fiquei - na linha 14;
garota - na linha 37;
```

```
mada - na linha 42;
na linha 6; - na linha 10;
na linha 22;
- na linha 22;
- na linha 25; - na linha 49; - na linha 62; - na linha 70;
na linha 25; - na linha 34; - na linha 39;
- na linha 66; - na linha 59; - na linha 67;
- na linha 46; - na linha 59; - na linha 67;
- na linha 2;
- na linha 3;
- na linha 32;
- na linha 36;
- na linha 36;
- na linha 36;
```