



Disciplina Algoritmos e Estruturas de Dados II	Curso	Turno Vespertino	Período 2º
Professores Gisele Pappa, Ítalo Cunha, William Schwartz			

Trabalho Prático Extra Codificação de Dados

Pontuação: **5 pontos**

Prazo de entrega: **05/07/2012**

Penalização por atraso: **redução de 5 pontos**

O objetivo deste trabalho é codificar caracteres de modo a reduzir o número de bits necessários para representar símbolos (caracteres da tabela ASCII).

O método de codificação a ser implementado consiste de duas etapas:

1. Geração de códigos de comprimento variável, o qual ordena decrescentemente de acordo com o número de ocorrências de cada símbolo (tabela 1) e agrupa os dois símbolos com menor probabilidade em um novo símbolo. Esta operação é repetida até que reste apenas um símbolo. A sequência de agrupamentos compõe uma árvore binária onde os símbolos de entrada estão nas folhas (figura 1).
2. Atribuição de um código a cada símbolo resultante. O código para cada símbolo é composto pelos dígitos obtidos no caminho a partir da raiz até o respectivo símbolo (nó folha), conforme ilustrado na tabela 2.

caractere	número de ocorrências
e	3320
h	1458
l	1067
o	1749
p	547
t	2474
w	266

Tabela 1: Número de ocorrências de cada símbolo.

caractere	dígitos do código resultante
e	00
h	011
l	110
o	010
p	1110
t	10
w	1111

Tabela 2: Código obtido para cada símbolo.

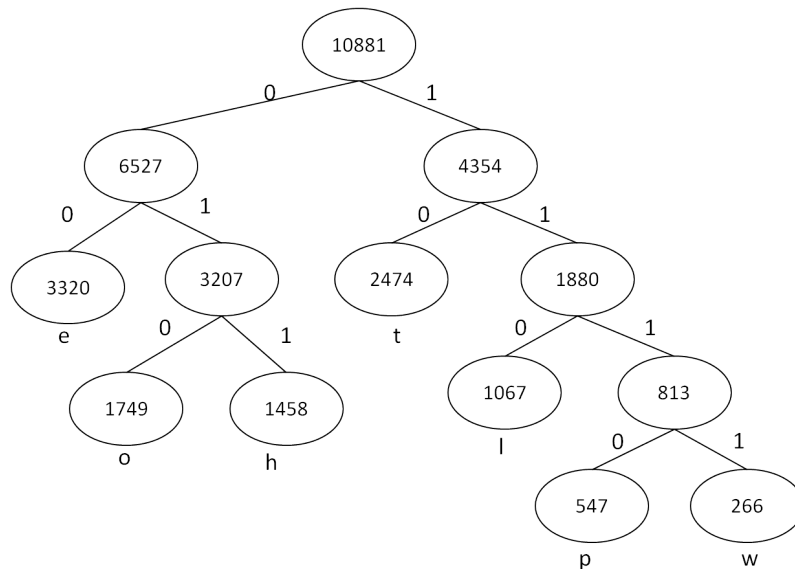


Figura 1: Árvore resultante da codificação.

O que deve ser feito

- Dado três arquivos de entrada:

1. arq1: arquivo de entrada contendo uma série de caracteres.
2. arq2: arquivo de entrada contendo uma série de caracteres.
3. arq3: arquivo de entrada contendo uma série de dígitos inteiros zeros e uns.

Implementar um programa em linguagem C capaz de:

1. Computar o número de ocorrências de cada caractere do arquivo arq1 (incluindo espaços e quebras de linha e ignorando caracteres acentuados) e gerar códigos utilizando o método de codificação descrito anteriormente.
2. Codificar o conteúdo do arquivo arq2 utilizando os códigos gerados no item 1. O resultado deve ser impresso no arquivo arq4 (apenas os dígitos devem ser impressos). Nota: os dígitos devem ser impressos como inteiros e não com codificação binária.
Exemplo: considerando o número de ocorrências conforme mostrado na tabela 1, resultando na codificação mostrada na tabela 2, tem-se:

```

arq2: pptwle
arq4: 1110111010111111000
  
```

3. Decodificar o conteúdo do arquivo arq3 utilizando os códigos gerados no item 1. O resultado deve ser impresso no arquivo arq5 (apenas os caracteres decodificados devem ser impressos).
Exemplo: considerando o número de ocorrências conforme mostrado na tabela 1, resultando na codificação mostrada na tabela 2, tem-se:

```

arq3: 000110111111110
arq5: ehht
  
```

- O seu programa deve ser executado da seguinte maneira:

```
./programa arq1.txt arq2.txt arq3.txt arq4.txt arq5.txt
```

O que deve ser entregue

- Deve ser submetido o código fonte, seus arquivos .h (com as declarações das funções) e seus arquivos .c (com a implementação das funções). O programa deve ser ANSI, ou seja, o código fonte do programa deve ser compilável no Linux, não devendo ser utilizada nenhuma biblioteca que seja específica do sistema operacional MS-Windows.
- Documentação sobre o trabalho (em .pdf), incluindo uma breve descrição das estruturas de dados utilizadas para resolver o problema e as decisões tomadas relativas aos casos e detalhes de especificação que porventura estejam omissos no enunciado. **A documentação deve ter no máximo uma página em fonte 11, espaçamento simples e com margens de 1,5 cm.**
- O trabalho será entregue através do sistema de submissão de trabalhos práticos: <http://aeds.dcc.ufmg.br>, em um arquivo compactado (.zip) contendo os arquivos fonte (arquivo .c, o arquivo .h, o main.c). A documentação deve ser enviada em pdf, no link especial para sua submissão no sistema.