

Практическая работа

Тема: Представление текстов. Сжатие информации

Цель: Практическое закрепление знаний о представлении текстовой информации в памяти компьютера.

Ход работы

Справочная информация

Текстовая информация состоит из символов: буквы, цифры, знаки препинания, скобки и др. Каждый символ имеет свой двоичный код, понятный для компьютера.

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера, называется *таблицей кодировки*.

Для разных типов ЭВМ используются разные таблицы кодировки. С распространением персональных компьютеров, международным стандартом стала таблица кодировки ASCII (American Standart Code for Information Interchange).

Таблица 1.1. Базовая таблица кодировки ASCII

32 пробел	48 0	64 @	80 P	96 `	112 p
33 !	49 1	65 A	81 Q	97 a	113 q
34 "	50 2	66 B	82 R	98 b	114 r
35 #	51 3	67 C	83 S	99 c	115 s
36 \$	52 4	68 D	84 T	100 d	116 t
37 %	53 5	69 E	85 U	101 e	117 u
38 &	54 6	70 F	86 V	102 f	118 v
39 '	55 7	71 G	87 W	103 g	119 w
40 (56 8	72 H	88 X	104 h	120 x
41)	57 9	73 I	89 Y	105 i	121 y
42 *	58 :	74 J	90 Z	106 j	122 z
43 +	59 ;	75 K	91 [107 k	123 {
44 ,	60 <	76 L	92 \	108 l	124
45 -	61 =	77 M	93]	109 m	125 }
46 .	62 >	78 N	94 ^	110 n	126 ~
47 /	63 ?	79 O	95 _	111 o	127

Задание 1. Декодировать текст, записанный в международной кодировочной таблице.

72 101 108 108 111 44 32 109 121 32 102 114 105 101 110 100 33

Задание 2. Используя таблицу кодировки ASCII, закодируйте сообщение:

Computing science

Справочная информация

Сжатием информации в памяти компьютера называют такое её преобразование, которое ведёт к сокращению объёма занимаемой памяти при сохранении закодированного содержания. Рассмотрим один из способов сжатия текстовой информации – алгоритм Хаффмена. С помощью этого алгоритма строится двоичное дерево, которое позволяет однозначно декодировать двоичный код, состоящий из символьных кодов различной длины. Двоичным называется дерево, из каждой вершины которого выходят две ветки. На рисунке приведен пример такого дерева, построенного для алфавита английского языка с учётом частоты встречаемости его букв.

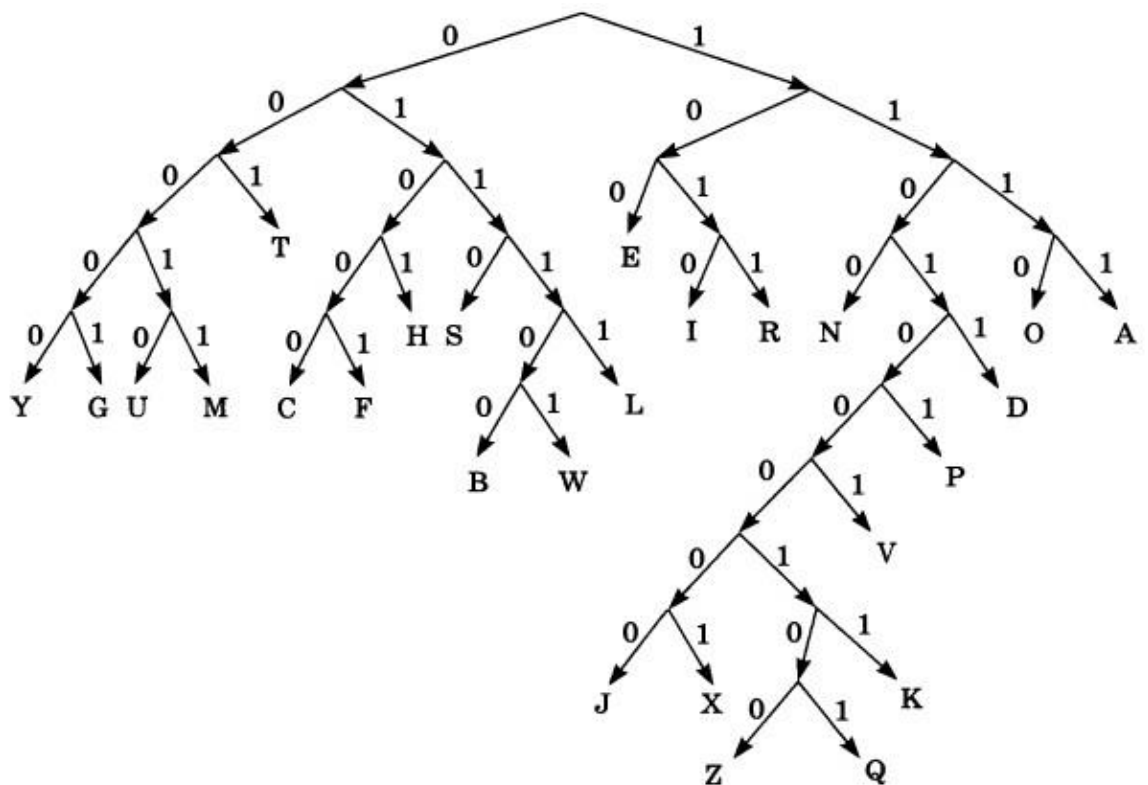


Рис. 1 Дерево Хаффмена

Закодируем с помощью данного дерева слово «hello»:

0101 100 01111 01111 1110

При размещении этого кода в памяти побитово он примет вид:

01011000 11110111 11110

Таким образом, текст, занимающий в кодировке ASCII 5 байтов, в кодировке Хаффмена займёт только 3 байта.

Задание 3. Используя метод сжатия Хаффмена, закодируйте следующие слова:

- 1) administrator
- 2) revolution
- 3) economy
- 4) department

Задание 4. Используя дерево Хаффмена, декодируйте следующие слова:

- 1) 01110011 1101001 10010110 10010111 100000
- 2) 00010110 01010110 10011001 0101101 01000100 000

Вывод