Задачи на кодирование текстовой информации и определение объема методика решения задач повышенного уровня сложности

С.Д. Богданова
ГБОУ средняя школа №314

Источник информации курсы повышения квалификации при СПБ ГУИТМО 2011г. "Теоретические основы информационных техногий.
(Методика подготовки школьников к олимпиадам по информатике)

2011

Что следует знать:

- ✔чаще всего используют кодировки, в которых на символ отводится 8 бит (8-битные) или 16 бит (16-битные)
- ✔необходимо запомнить, что

  1 байт = 8 бит
  1 Кбайт = 1024 байтам
  1 Мбайт = 1024 Кбайтам
  1 Гбайт = 1024 Мбайтам
  1 Тбайт = 1024 Гбайтам
- ✔чтобы найти информационный объем текста I, нужно умножить количество символов N на число бит, приходящих на один символ K

  $$I = N \cdot K$$
- ✔после знака препинания внутри текста ставится пробел
- ✔все символы кодируются одинаковым числом бит
- ✔если текст занимает 2 строки, то в конце 1-й, вместо пробела, стоит невидимый знак переноса, т.е. +1 символ, если строк -3, то +2 знака и т.д.

Задачи на кодирование текстовой информации и определение объема

1. Считая, что каждый символ кодирование одним байтом, оцените информационный объём предложения:

«Ты – душная как май, Ямская,

Шевардина ночной редут,»

Ответ: 54 байта

2. При шифровании текста использовался 6-битовый метод кодирования символов. Оцените объём следующего предложения в этой кодировке.

не тот друг, кто жалеет, а тот, кто помогает.

Ответ: 270 бит

3. Дано 16 слов, состоящих из 6 символов. Если каждый символ кодируется 1 байтом, то какое минимальное количество бит понадобится для кодирования всех слов?

Ответ: 768 бит

Задачи на кодирование текстовой информации и определение объема

1. Мощность алфавита равна 64. Сколько Кб памяти потребуется, чтобы сохранить 128 страниц текста, содержащего в среднем 256 символов на каждой страницу?

Решение

1) $N = 64$ $2^i = 64$ $i=6$ бит

2) $256 \cdot 128 = 32\ 768$ символов

3) $32\ 768 \cdot 6 = 196\ 608$ бит

196 608 бит = 24 576 байт = 24 Кб

Задачи на кодирование текстовой информации и определение объема

2. Известно, что на каждой странице документа 128 строк, содержащих по 48 символов каждая. Сколько страниц в документе, если его информационный объём при условии, что каждый символ кодировался 2-байтовой кодировкой Unicode, составил 720 Кбайт?

Решение

1) $128 \cdot 48 = 6\ 144$ символа

2) $6\ 144 \cdot 2 = 12\ 288$ байт

I = 720 Кб = 737 280 байт

3) $737\ 280 : 12\ 288 = 60$ страниц

3. Объём сообщения, содержащего 1 024 символа, составил 1/512 часть Мб. Каков размер алфавита, с помощью которого записано сообщение?

Решение

I= 1/512 Мб = 2 048 байт

1) $I=K\cdot i$ $2\,048=1\,024\cdot i$ i=2байта=16 бит
2) $N=2^i$ $2^{16}=65\,536$ символов

4. Дан текст из 1 000 символов. Известно, что символы берутся из таблицы размером 16x32. Определить объём текста в битах.

Решение

1) N= 16x32 = 512 символов
2) $N=2^i$ $512=2^i$ i=9 бит
3) $I=K\cdot i$ $I=1000\cdot 9= 9\,000$ бит

Задачи на кодирование текстовой информации и определение объема

1. Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке длиной в 20 символов, первоначально записанного в 2-байтном коде Unicode, в 8-битную кодировку КОИ-8. На сколько бит уменьшилась длина сообщения?

Решение

1) при 16-битной кодировке объем сообщения = 16 · 20 бит
2) когда его перекодировали в 8-битный код, его объем стал = 8 · 20 бит
3) таким образом, сообщение уменьшилось на

$$16 \cdot 20 - 8 \cdot 20 = 320 - 160 = 160 \text{ бит}$$

или

$$16 \cdot 20 - x = 8 \cdot 20$$

$$320 - x = 160$$

$$x = 160 \text{ бит} = 20 \text{ байт}$$

Задачи на кодирование текстовой информации и определение объема

2. Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде Unicode, в 8-битную кодировку КОИ-8. При этом информационное сообщение уменьшилось на 480 бит. Какова длина сообщения в символах?

Решение

1) Обозначим через **x** количество символов

2) Тогда при 16-битной кодировке объем сообщения

$$16 \cdot x \text{ бит}$$

3) Когда его перекодировали в 8-битный код, его объем стал равен

$$8 \cdot x \text{ бит}$$

4) Зная, что информационное сообщение уменьшилось на 480 бит, имеем

$$x \cdot 16 - 480 = x \cdot 8$$

$$x \cdot 8 = 480$$

$$x = 60 \text{ символов}$$

Задачи на кодирование текстовой информации и определение объема

1. В некоторой стране автомобильный номер состоит из 7 символов. В качестве символов используют 18 различных букв и десятичные цифры в любом порядке. Каждый такой номер в компьютерной программе записывается минимально возможным и одинаковым целым количеством байтов, при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов. Определите объем памяти, отводимый этой программой для записи 60 номеров.

Решение

1) Найдём сколько всего используется символов

18+10=28 символов

2) Тогда 1 символ кодируется 5 битами ($2^4<28<2^5$)

3) Отсюда 1 номер занимает объем в памяти

5 бит·7 символов=35 бит ≈ 5 байт (т.к. байт только целое число)

4) Объем памяти = 5 байт·60 номеров = 300 байт

Задачи на кодирование текстовой информации и определение объема

2. Для регистрации на сайте некоторой страны пользователю требуется придумать пароль. Длина пароля – ровно 11 символов. В качестве символов используются десятичные цифры и 12 различных букв местного алфавита, причём все буквы используются в двух начертаниях: как строчные, так и заглавные (регистр буквы имеет значение). Под хранение каждого такого пароля на компьютере отводится минимально возможное и одинаковое целое количество байтов, при этом используется посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов. Определите объём памяти, который хранение 60 паролей.

Решение

1) Найдём сколько всего используется символов

$$10+12\cdot 2=34 \text{ символа}$$

2) Тогда 1 символ кодируется 6 битами ($2^5<34<2^6$)

3) Отсюда 1 символ занимает объем в памяти

$$6 \text{ бит}\cdot 11 \text{ символов}=66 \text{ бит} \approx 9 \text{ байт}$$

4) Объем памяти = 9 байт·60 паролей = 540 байт

3. Репетиционный экзамен в школе сдают 125 человек. Каждый из них выделяют специальный номер, идентифицирующий его в автоматической системе проверки ответов. При регистрации участника для записи его номера система использует минимально возможное количество бит, одинаковое для каждого участника. Каков объём информации в битах, записанный устройством после регистрации 60 участников?

Решение

1) Найдём сколько бит потребуется для кодирования 125 человек

7 бит $(2^6<125<2^7)$

2) 7 бит · 60 участников = 420 бит

Задачи на кодирование текстовой информации и определение объема

1. Для кодирования букв А, Б, В, Г решили использовать двухразрядные последовательные двоичные числа (от 00 до 11, соответственно). Если таким способом закодировать последовательность символов БАВГ и записать результат шестнадцатеричным кодом, что получится.

Решение

1) из условия имеем - А – 00, Б – 01, В – 10 и Г – 11, код равномерный
2) последовательность БАВГ кодируется так - 01 00 10 11 = 1001011
3) разбиваем такую запись на тетрады справа налево и каждую тетраду переведем в шестнадцатеричную систему (то есть, сначала в десятичную, а потом заменим все числа от 10 до 15 на буквы A, B, C, D, E, F); получаем

$$1001011 = 0100\ \ 1011_2 = 4B_{16}$$

2. Для 5 букв латинского алфавита заданы их двоичные коды (для некоторых букв - из двух бит, для некоторых - из трех). Эти коды представлены в таблице:

A	B	C	D	E
000	01	100	10	011

Определить, какой набор букв закодирован двоичной строкой 0110100011000

1) EBCEA 2) BDDEA 3) BDCEA 4) EBAEA

Решение

в данном случае самое простое и надежное – просто закодировать все ответы, используя приведенную таблицу кодов, а затем сравнить результаты с заданной цепочкой

получим

1) EBCEA – 011 01 100 011 000 2) BDDEA – 01 10 10 011 000

3) BDCEA – 01 10 100 011 000 4) EBAEA – 011 01 000 011 000

сравнивая эти цепочки с заданной, находим, что правильный ответ - 3

3. В таблице ниже представлена часть кодовой таблицы ASCII:

Символ	1	5	A	B	Q	a	b
Десятичный код	49	53	65	66	81	97	98
Шестнадцатеричный код	31	35	41	42	51	61	62

Каков шестнадцатеричный код символа «q» ?

Решение

1) разница кодов букв «q» и «a» равна разнице кодов букв «Q» и «A», то есть, $51_{16} - 41_{16} = 10_{16}$
2) тогда шестнадцатеричный код символа «q» равен коду буквы «a» + 10_{16}
3) отсюда находим

$$61_{16} + 10_{16} = 71_{16}.$$

Задачи на кодирование текстовой информации и определение объема

Лазерный принтер Canon LBP печатает со скоростью в среднем 6,3 Кбит в секунду. Сколько времени понадобится для распечатки 8-ми страничного документа, если известно, что на одной странице в среднем по 45 строк, в строке 70 символов

Решение

1) Найдём количество информации, содержащейся на 1 странице:

45 ·70 = 3 150 байт · 8 бит = 25 200 бит

2) Затем количество информации на 8 страницах:

25 200·8 = 201 600 бит

3) Переведём 6,3 Кбит/сек

I=6,3Кб=6 451,2 бит/сек.

4) Находим время печати: 201600: 6451,2 = 31,25 секунд.