

ЕГЭ

Урок 4

**Кодирование
текстовой информации**

Двоичное кодирование текстовой информации в компьютере

Для представления текстовой информации (прописные и строчные буквы русского и латинского алфавитов, цифры, знаки и математические символы) достаточно **256** различных знаков.

По формуле

$$N = 2^I$$

можно вычислить, какое количество информации необходимо, чтобы закодировать каждый знак:

$$N = 2^I \Rightarrow 256 = 2^I \Rightarrow 2^8 = 2^I \Rightarrow \underline{I = 8 \text{ битов}}$$

Для обработки текстовой информации на компьютере необходимо представить ее в двоичной знаковой системе.

Каждому знаку необходимо поставить в соответствие уникальный 8-битовый двоичный код, значения которого находятся в интервале от 00000000 до 11111111 (в десятичном коде от 0 до 255).

Присвоение знаку конкретного двоичного кода — это вопрос соглашения, которое фиксируется в кодовой таблице.

К сожалению, в настоящее время существуют пять различных кодовых таблиц для русских букв (Windows, MS-DOS, КОИ-8, Mac, ISO), поэтому тексты, созданные в одной кодировке, не будут правильно отображаться в другой.

В последние годы широкое распространение получил новый международный стандарт кодирования текстовых символов **Unicode**, который отводит на каждый символ **2 байта (16 битов)**.

По формуле

$$N = 2^I$$

можно определить количество символов, которые можно закодировать:

$$N = 2^I = 2^{16} = \underline{65\ 536}$$

Такого количества символов оказалось достаточно, чтобы закодировать не только русский и латинский алфавиты, цифры, знаки и математические символы, но и греческий, арабский, иврит и другие алфавиты.

Windows, MS-DOS, КОИ-8, Mac, ISO – 8-битовый двоичный код – 256 символов

Unicode – 16-битовый двоичный код – 65 536 символов

Задачи

1. Во сколько раз увеличится информационный объем страницы текста (текст не содержит управляющих символов форматирования) при его преобразовании из кодировки Windows (таблица кодировки содержит **256** символов) в кодировку Unicode (таблица кодировки содержит **65 536** символов)?

1) в **2** раза 2) в 8 раз 3) в 16 раз 4) в 256 раз

Решение:

Количество информации, необходимое, чтобы закодировать каждый знак в кодировке Windows (256 символов):

$$N = 2^I \Rightarrow 256 = 2^I \Rightarrow 2^8 = 2^I \Rightarrow \underline{I = 8 \text{ битов.}}$$

Количество информации, необходимое, чтобы закодировать каждый знак в кодировке Unicode (65 536 символов):

$$N = 2^I \Rightarrow 65\,536 = 2^I \Rightarrow 2^{16} = 2^I \Rightarrow \underline{I = 16 \text{ битов.}}$$

Для кодировки каждого знака необходимо в два раза больше информации, следовательно информационный объем страницы текста увеличится **в 2 раза**.

2. Какое количество информации необходимо для кодирования каждого из 65 536 символов алфавита?

- 1) 1 байт 2) 2 байта 3) 8 битов 4) 32 бита

Решение:

Количество информации, необходимое, чтобы закодировать каждый знак в кодировке Unicode (65 536 символов):

$$N = 2^I \Rightarrow 65\,536 = 2^I \Rightarrow 2^{16} = 2^I \Rightarrow I = 16 \text{ битов} = = \underline{2 \text{ байта}}$$

3. В таблице кодов ASCII имеют международный стандарт:

- 1) первые 64 кода 2) первые 128 кодов
3) последние 128 кодов 4) первые 127 кодов

(0 – 32) – коды операций

(33 – 127) – интернациональные (международные) коды

(128 – 255) – национальные (русские буквы)

4. Скорость передачи данных через модемное соединение равна 32 Кбит/с. Передача текстового файла через это соединение заняла 15 с. Определите, сколько страниц содержал переданный текст, если известно, что он был представлен в кодировке Unicode, а на одной странице – 48 символов.
- 1) 10 2) 640 3) 1280 4) 10240

Решение:

Объем текстового файла:

$$32 \text{ Кбит/с} * 15 \text{ с} = 480 \text{ Кбит}$$

Объем одной страницы (Unicode – 1 символ – 16 бит):

$$48 \text{ символов} * 16 \text{ бит} = 768 \text{ бит}$$

Количество страниц:

$$480 \text{ Кбит} = 480 * 1024 = 491\,520 \text{ бит}$$

$$491\,520 \text{ бит} / 768 \text{ бит} = 640 \text{ (страниц)}$$

Ответ: 640

5. 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?

1) 30

2) 60

3) 120

4) 480

Решение:

x – длина сообщения в символах

$16 * x$ бит – объем сообщения в 16-битном коде Unicode

$8 * x$ бит – объем сообщения в 8-битной кодировке КОИ-8

$$16x - 8x = 480$$

$$8x = 480$$

$$x = 60$$

Ответ: 60

6. Система оптического распознавания символов позволяет преобразовывать отсканированные изображения страниц документа в текстовый формат со скоростью 4 страницы в минуту и использует алфавит мощностью 65 536 символов.

Какое количество информации будет нести текстовый документ, каждая страница которого содержит 40 строк по 50 символов, после 10 минут работы приложения?

Решение:

Информационная емкость 1 символа алфавита:

$$N = 2^I, \quad 65\,536 = 2^I, \quad 2^{16} = 2^I, \quad \underline{I = 16 \text{ битов}}$$

Количество информации на странице:

$$16 \text{ битов} * 40 * 50 = 32\,000 \text{ битов}$$

$$1 \text{ байт} = 8 \text{ бит}, \quad 32\,000 \text{ битов} / 8 = 4\,000 \text{ байтов}$$

Количество информации, которое будет нести текстовый документ:

$$4\,000 \text{ байтов} * 4 * 10 = 160\,000 \text{ байтов} \approx \underline{156 \text{ Кбайт}}$$

Ответ: 156 Кбайт

Домашняя работа №4 (ЕГЭ!)

Кодирование текстовой информации (11 задач)

1. Во сколько раз уменьшится информационный объем страницы текста (текст не содержит управляющих символов форматирования) при его преобразовании из кодировки Unicode (таблица кодировки содержит 65 536 символов) в кодировку Windows (таблица кодировки содержит 256 символов)?
2. Два текста содержат одинаковое количество символов. Первый текст составлен в алфавите мощностью 16 символов. Второй текст в алфавите мощностью 256 символов. Во сколько раз количество информации во втором тексте больше, чем в первом?
1) 2 2) 4 3) 12 4) 24

3. Какое количество информации необходимо для кодирования каждого из 256 символов алфавита?

4. Для передачи секретного сообщения используется код, состоящий из десятичных цифр. При этом все цифры кодируются одним и тем же (минимально возможным) количеством бит. Определите информационный объем сообщения длиной в 150 символов.

5. 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 640 бит.

Какова длина сообщения в символах?

1) 30

2) 60

3) 80

4) 480

6. Средняя скорость передачи данных с помощью модема равна 30 Кбит/с. Определите, сколько секунд понадобится модему, чтобы передать 80 страниц текста в кодировке КОИ-8, если считать, что на каждой странице в среднем 96 символа?

7. В какой кодовой таблице можно закодировать 65536 различных символов?

- 1) КОИ-8 2) CP1251 3) ASCII 4) Unicode

8. В настоящее время существует пять основных кодировок кириллицы. Какая из перечисленных кодировок не относится к их числу?

- 1) RADIX-50 2) KOI8-R 3) CP1251 4) ISO 8859-5

9. Какие символы в таблице ASCII могут быть зашифрованы десятичными кодами 87 и 136?

- 1) D и W 2) W и И 3) Б и Я 4) Б и b

10. В кодировке Unicode на каждый символ отводится два байта. Определите информационный объем слова из двадцати четырех символов в этой кодировке.

- 1) 384 бита 2) 192 бита 3) 256 бит 4) 48 бит

11. Используется кодовая таблица CP1251 (Windows Cyrillic). Сколько килобайт будет занимать файл в простом текстовом формате (plain text), если в тексте 200 страниц, на странице 32 строки, а в строке в среднем 48 символов?

- 1) 307,2 2) 300 3) 384 4) 2 400