
Тема: Кодирование текстовой информации. Кодировка ASCII.

Основные кодировки кириллицы.

Что нужно знать:

- все символы кодируются одинаковым числом бит (алфавитный подход)
- чаще всего используют кодировки, в которых на символ отводится 8 бит (8-битные) или 16 бит (16-битные)
- при измерении количества информации принимается, что в одном байте 8 бит, а в одном килобайте (1 Кбайт) – 1024 байта, в мегабайте (1 Мбайт) – 1024 Кбайта
- после знака препинания внутри (не в конце!) текста ставится пробел
- чтобы найти информационный объем текста I , нужно умножить количество символов N на число бит на символ K : $I = N \cdot K$
- две строчки текста не могут занимать 100 Кбайт в памяти

Пример задания:

Определите информационный объем текста

Бамбарбия! Кергуду!

- 1) 38 бит 2) 144 бита 3) 152 бита 4) 19 бит

Решение:

- 1) в этом тексте 19 символов (обязательно считать пробелы и знаки препинания)
- 2) если не дополнительной информации, считаем, что используется 8-битная кодировка (чаще всего явно указано, что кодировка 8- или 16-битная)
- 3) поэтому в сообщении $19 \cdot 8 = 152$ бита информации (ответ 3).

Возможные ловушки:

- указано правильное число, но другие единицы измерения (объем текста 19 *байт*, а один из неверных ответов – 19 *бит*)
- расчет на то, что «забудут» пробел, в этом случае получается $18 \cdot 8 = 144$ бита (ответ 2, неверный)
- в 16-битной кодировке объем текста – 38 *байт*, а один из неверных ответов – 38 *бит*.

Еще пример задания:

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:

- 1) обозначим количество символов через N
- 2) при 16-битной кодировке объем сообщения – $16 \cdot N$ бит
- 3) когда его перекодировали в 8-битный код, его объем стал равен – $8 \cdot N$ бит
- 4) таким образом, сообщение уменьшилось на $16 \cdot N - 8 \cdot N = 8 \cdot N = 480$ бит
- 5) отсюда находим $N = 480 / 8 = 60$ символов (ответ 2).

Еще пример задания:

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» ?

1) 71_{16}

2) 83_{16}

3) $A1_{16}$

4) $B3_{16}$

Решение:

- 1) в кодовой таблице ASCII (*American Standard Code for Information Interchange, американский стандартный код для обмена информацией*) все заглавные латинские буквы A-Z расставлены по алфавиту, начиная с символа с кодом $65=41_{16}$
- 2) все строчные латинские буквы a-z расставлены по алфавиту, начиная с символа с кодом $97=61_{16}$
- 3) отсюда следует, что разница кодов букв «q» и «a» равна разнице кодов букв «Q» и «A», то есть, $51_{16} - 41_{16} = 10_{16}$
- 4) тогда шестнадцатеричный код символа «q» равен коду буквы «a» плюс 10_{16}
- 5) отсюда находим $61_{16} + 10_{16} = 71_{16}$ (ответ 1).

Задачи для тренировки:

- 1) Считая, что каждый символ кодируется одним байтом, определите, чему равен информационный объем следующего высказывания *Жан-Жака Руссо*:
Тысячи путей ведут к заблуждению, к истине – только один.
1) 92 бита 2) 220 бит 3) 456 бит 4) 512 бит
- 2) Считая, что каждый символ кодируется одним байтом, определите, чему равен информационный объем следующего высказывания *Алексея Толстого*:
Не ошибается тот, кто ничего не делает, хотя это и есть его основная ошибка.
1) 512 бит 2) 608 бит 3) 8 Кбайт 4) 123 байта
- 3) Считая, что каждый символ кодируется одним байтом, определите, чему равен информационный объем следующего высказывания *Рене Декарта*:
Я мыслю, следовательно, существую.
1) 28 бит 2) 272 бита 3) 32 Кбайта 4) 34 бита
- 4) 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
- 5) Считая, что каждый символ кодируется 16-ю битами, оцените информационный объем следующей пушкинской фразы в кодировке *Unicode*:
Привычка свыше нам дана: Замена счастию она.
1) 44 бита 2) 704 бита 3) 44 байта 4) 704 байта
- 6) Считая, что каждый символ кодируется одним байтом, оцените информационный объем следующего предложения из пушкинского четверостишия:
Певец-Давид был ростом мал, Но повалил же Голиафа!
1) 400 бит 2) 50 бит 3) 400 байт 4) 5 байт
- 7) Считая, что каждый символ кодируется одним байтом, оцените информационный объем следующего предложения:
Мой дядя самых честных правил, Когда не в шутку занемог, Он уважать себя заставил И лучше выдумать не мог.
1) 106 бит 2) 848 бит 3) 106 Кбайт 4) 848 Кбайт
- 8) Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 8-битном коде, в 16-битную кодировку *Unicode*. При этом информационное сообщение увеличилось на 2048 байт. Каков был информационный объем сообщения до перекодировки?
1) 1024 байт 2) 2048 бит 3) 2 Кбайта 4) 2 Мбайта
- 9) Считая, что каждый символ кодируется 16-ю битами, оцените информационный объем следующей фразы в кодировке *Unicode*:
В шести литрах 6000 миллилитров.

-
- 1) 1024 байта 2) 1024 бита 3) 512 байт 4) 512 бит

10) Считая, что каждый символ кодируется 16-ю битами, оцените информационный объем следующего предложения:

Блажен, кто верует, тепло ему на свете!

- 1) 78 бит 2) 80 байт 3) 312 бит 4) 624 бита

11) Считая, что каждый символ кодируется одним байтом, оцените информационный объем следующего предложения:

Белеет Парус Одинокий В Тумане Моря Голубом!

- 1) 352 бита 2) 44 бита 3) 352 байта 4) 88 байт

12) 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 800 бит. Какова длина сообщения в символах?

- 1) 50 2) 100 3) 200 4) 800

13) В таблице ниже представлена часть кодовой таблицы ASCII:

Символ	1	5	Ж	К	Р	j	k
Десятичный код	49	53	74	75	80	106	107
Шестнадцатеричный код	31	35	4A	4B	50	6A	6B

Каков шестнадцатеричный код символа «р» ?

- 1) 70_{16} 2) 85_{16} 3) $6F_{16}$ 4) CE_{16}

14) В таблице ниже представлена часть кодовой таблицы:

Символ	С	Т	У	Я	с	т	у
Десятичный код	145	146	147	159	225	226	227
Шестнадцатеричный код	91	92	93	9F	E1	E2	E3

Каков шестнадцатеричный код символа «я» ?

- 1) $2A_{16}$ 2) DF_{16} 3) EF_{16} 4) 180_{16}