



Задание №11, Кодирование

Теория

Алфавит — множество (=набор) символов.

Мощность алфавита — количество символов в алфавите.

Объём символа — то, сколько памяти занимает один символ.

Объём сообщения — то, сколько памяти занимает сообщение.

Под объёмом в данном задании подразумевается информационный объём, в двоичном виде



Объём измеряется в единицах хранения цифровой информации:

- Бит — единица, принимающая значение 1 или 0, один разряд двоичного кода
- Байт — 8 бит (2^3 бит)
- Килобайт (Кб) — $2^{10} = 1024$ байт (2^{13} бит)
- Мегабайт (Мб) — 1024 Кб (2^{23} бит)
- Гигабайт (Гб) — 1024 Мб (2^{33} бит)
- Терабайт (Тб) — 1024 Гб (2^{43} бит)

Формула для определения мощности алфавита:

$$N = 2^i$$

Формула для определения объема сообщения:

$$V = n \cdot i, \text{ где}$$

V — объем сообщения;

n — количество символов;

i — объем символа. (в битах)



Очень советую запомнить значения степеней двойки хотя бы до 10, часто встречаются в информатике и в ЕГЭ в частности, особенно в этом задании:

Степени числа 2

2^0	1	2^8	256
2^1	2	2^9	512
2^2	4	2^{10}	1024
2^3	8	2^{11}	2048
2^4	16	2^{12}	4096
2^5	32	2^{13}	8192
2^6	64	2^{14}	16384
2^7	128	2^{15}	32768

11 задание ЕГЭ считается относительно простым, но в нём встречаются много разных формулировок и условий, поэтому важно понимать основные обозначения и вчитываться в условие, а не следовать шаблону.

Часто встречающиеся формулировки:

- "В алфавите строчные и заглавные буквы" \Rightarrow умножение количества символов алфавита на 2 (без учёта цифр).
- "Кодируются минимальным возможным количеством **байт**" \Rightarrow делим полученный информационный объём в битах на 8, округляем всегда в

большую сторону, т.е. 17 бит это уже 3 байта.

- Если в условии даётся какой-то "лимит", например, "на 10 пользователей потребовалось 32 байта", то объём информации на одного пользователя **округляем в меньшую сторону**: 3 байта.
- Разные типы кодирования для разных данных:
 - Посимвольное, каждый элемент кодируется отдельно, подходит для строк: каждый символ весит i , где мощность алфавита $N=2^i$. Общий объём соответственно $i * n$, где n - количество символов в строке.
 - "Целое", каждый вариант кодируется уникально, подходит для чисел, дат и подобного: алфавит - набор всех возможных вариантов. Общий объём равен i .

Максимальная мощность алфавита: 2^i

Минимальная мощность алфавита: $2^{(i-1)} + 1$



Максимальная мощность алфавита при $i = 10$ равна 1024, т.к. количество различных комбинаций десятиразрядного двоичного числа (которое и будет весить 10 бит) равно 2 в 10 степени, а значит равное такое количество вариантов и можно закодировать. Минимальная мощность алфавита в свою очередь для $i = 10$ равна 513, т.к. 512 можно закодировать имея $i=9$, а 513-ому варианту уже не хватит уникальной комбинации для кодирования, поэтому и придётся брать $i = 10$.

Алгоритм решения

Типичная задача:

(разбиралась на занятии)

1 При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 16 символов. 2 В качестве символов используют прописные и строчные буквы латинского алфавита, а также десятичные цифры. 3 В базе данных для хранения сведений о каждом пользователе отведено одинаковое и минимально возможное целое число байт. 4 При этом используют посимвольное кодирование паролей, все символы кодируют одинаковым и минимально возможным количеством бит. 5 Кроме собственно пароля, для каждого пользователя в системе хранятся дополнительные сведения, для чего выделено 20 байт на одного пользователя. 6 В компьютерной системе отведено 10 Кбайт для хранения сведений о пользователях. 7 О каком наибольшем количестве пользователей может быть сохранена информация в системе? 8 В ответе запишите только целое число – количество пользователей.

1. В первом же предложении видим длину пароля. Обозначим количество символов в нём за **$n = 16$** .
2. Второе предложение задаёт нам количество символов в алфавите (его мощность) для пароля: **$N = 26$ (количество букв в латинском алфавите) * 2 (прописные + строчные) + 10 (цифры) = 62**.
3. Учитывая 4 предложение, находим объём (цифровой вес) одного символа: **$i = \log_2(N) = 6$ (округляем всегда в большую сторону)**.
4. Объём пароля равен **$V(\text{password}) = n * i = 16 * 6 = 96$ бит**.
5. Из 5 предложения понимаем, что у каждого пользователя также хранятся дополнительные сведения объёмом **$V(\text{additional}) = 20$ байт**. Общий объём данных о пользователе равен **$V(\text{user}) = V(\text{password}) + V(\text{additional}) = 20 \text{ байт} + 96 \text{ бит} = 20 * 8 + 96 = 256$ бит**.
6. Из 3 предложения узнаём, что для данных о пользователе выделено целое число байт. Тогда **$V(\text{user}) = 256 \text{ бит} = 256/8 \text{ байт} = 32 \text{ байта}$ (если бы биты не делились нацело, мы бы округляли в большую сторону)**.
7. Находим ответ, деля выделенные системе 10 Кбайт на объём данных одного пользователя (согласно 3 предложению каждому пользователю выделено одинаковое количество байт):

$$10 * 1024 \text{ (переводим Кб в байты)} / 32 = \underline{320}.$$

Обратная задача:

(разбиралась на занятии)

При регистрации в компьютерной системе каждому пользователю выдаётся идентификатор, состоящий из 33 символов и содержащий только символы из 18-буквенного набора. В базе данных для хранения сведений о каждом пользователе отведено одинаковое целое число байт, при этом для хранения сведений о 768 пользователях используется 21 Кбайт. Для каждого пользователя хранятся идентификатор и дополнительные сведения. Каждый символ в идентификаторе кодируется одинаковым и минимально возможным количеством бит. На хранение идентификатора отведено минимальное возможное целое количество байт. Сколько байт отведено для хранения дополнительных сведений о пользователе?

1. Аналогично предыдущей задаче, находим информационный объём идентификатора:

$n = 33; N = 18 \Rightarrow i = 5; V(id) = 33 * 5 = 165 \text{ бит} \Rightarrow$ т.к. в предпоследнем предложении условия указано, что на хранение идентификатора выделяется минимальное целое число байт, то переводим биты в байты (делим на 8) с округлением в большую сторону **$\Rightarrow V(id) = 21 \text{ байт}$.**

2. Находим, сколько байт выделяется для хранения информации об одном пользователе:

$V(user) = 21 * 1024 / 768 = 28 \text{ байт}$.

3. Находим, сколько байт остаётся для хранения дополнительных данных о пользователе:

$V(additional) = V(user) - V(id) = 28 - 21 = \underline{7 \text{ байт}}$.

Усложнённый вариант:

(разбирался на занятии)

При регистрации на государственном портале пользователя просят ввести следующие данные: дата рождения, номер паспорта и адрес проживания. Дата рождения состоит из дня (1 - 31), месяца (1 - 12) и года (1800 - 2400), при этом для хранения даты отводится битовая последовательность одинаковой минимальной длины для всех пользователей, которая представляет собой одно двоичное число. Номер паспорта представлен как строка из 10 цифр от 0 до 9, каждая из которых кодируется одинаковым и минимально возможным количеством бит. Известно, что для кодирования информации об одном пользователе выделяется целое, одинаковое для всех пользователей минимальное количество байт. Известно, что адрес проживания содержит символы из алфавита, состоящего из 33 символов, при этом используется посимвольное кодирование, и каждый символ кодируется одинаковым и минимально возможным количеством бит, для записи адреса проживания используются прописные буквы.

Известно, что для хранения данных о 1222 пользователях понадобилось 36 Кбайт памяти. Найдите максимальную длину строки, которая может быть адресом пользователя.

1. Т.к. в условии не указано, что для кодирования даты рождения используется посимвольное кодирование, то кодируется каждый возможный вариант целиком (т.е. 01.01.1800 кодируется не как совокупность кодов $0 = 0...00$, $1 = 0...01$, $8 = 0...11$, а как в целом $01.01.1800 = 0...00$, $02.01.1800 = 0...01$ и т.д.). Тогда мощностью алфавита будет количество всех возможных вариантов дат: $N(\text{date}) = 31 * 12 * 601 = 223\,572 \Rightarrow i(\text{date}) = 18$ ($2^{18} = 262\,144$). В этом случае $V(\text{date}) = i(\text{date}) = 18$ бит.
2. Номер паспорта кодируется посимвольно, т.к. в условии указано, что закодированы именно цифры: $n(\text{passport}) = 10$; $N(\text{passport}) = 10 \Rightarrow i(\text{passport}) = 4$; $V(\text{passport}) = 10 * 4 = 40$ бит.
3. Адрес кодируется посимвольно, прямо указывается в условии: $N(\text{address}) = 33 \Rightarrow i(\text{address}) = 6$ бит.
4. Находим, сколько памяти выделяется для одного пользователя, учитывая, что в условии сказано, что выделяется целое одинаковое минимальное количество байт: $V(\text{user}) = 36 * 1024 / 1222 = 30,1669..$ байт $\Rightarrow 30$ байт (тут округляем в меньшую сторону, т.к. если взять 31 байт, то общий объём перевалит за 36 Кб)
5. Про дату рождения, номер паспорта и адрес не сказано, что они хранятся целым количеством байт, как в прошлых задачах, поэтому оставляем их в битах. Чтобы найти максимальную длину строки адреса, нужно вычесть

из общего выделенного объёма для пользователя фиксированные объёмы даты и номера паспорта и поделить на объём адресного объёма символа:
 $n(\text{address}) = (V(\text{user}) - V(\text{date}) - V(\text{id})) / i(\text{address}) = (30 * 8$ (переводим в биты) - 40 - 18) / 6 = 182 / 6 = 30 (округляем в меньшую сторону, т.к. 31-ый символ не поместится в оставшийся свободный альбом).

6. Ответ 30