# 13 (повышенный уровень, время – 3 мин)

**Тема**: Вычисление информационного объема сообщения.

**Что нужно знать**:

* с помощью *K* бит можно закодировать различных вариантов (чисел)
* таблица степеней двойки, она же показывает, сколько вариантов *Q* можно закодировать с помощью *K*  бит:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *K*, бит | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *Q*, вариантов | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |

* 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[[1]](#footnote-1)
* чтобы найти информационный объем сообщения (текста) *I*, нужно умножить количество символов (отсчетов) *N* на число бит на символ (отсчет) *K*: 
* две строчки текста не могут занимать 100 Кбайт в памяти
* мощность алфавита *M*  – это количество символов в этом алфавите
* если алфавит имеет мощность *M*, то количество всех возможных «слов» (символьных цепочек) длиной *N* (без учета смысла) равно ; для двоичного кодирования (мощность алфавита *M* – 2 символа) получаем известную формулу: 

### Пример задания:

**Р-08.** При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 11 символов. Из соображений информационной безопасности каждый пароль должен содержать хотя бы 2 десятичных цифры, как прописные, так и строчные латинские буквы, а также не менее 2-х символов из 6-символьного набора: «&», «#», «$», «\*», «!», «@». В базе данных для хранения сведений о каждом пользователе отведено одинаковое и минимально возможное целое число байт. При этом используют посимвольное кодирование паролей, все символы кодируют одинаковым и минимально возможным количеством бит. Кроме собственно

пароля, для каждого пользователя в системе хранятся дополнительные сведения, для чего выделено целое число байт; это число одно и то же для всех пользователей. Для хранения сведений о 30 пользователях потребовалось 900 байт. Сколько байт выделено для хранения дополнительных сведений об одном пользователе? В ответе запишите только целое число – количество байт.

**Решение**:

1. если бы мы знали точно, сколько цифр и сколько специальных символов содержит пароль и где точно они расположены, можно было бы использовать «раздельное» кодирование: на кодирование цифр использовать по 4 бита (24 > 10), на кодирование спецсимволов – по 3 бита (23 > 6), а на кодирование остальных символов (латинских букв) – по 6 бит (26 > 26·2=52)
2. поскольку количество и месторасположение цифр и спецсимволов а пароле неизвестно, нужно рассматривать полный набор символов: 10 + 6 + 26·2 = 68
3. при этом на каждый символ нужно выделить 7 бит (27 > 68)
4. на 11 символов пароля выделяется 77 бит, округляя вверх до целого числа байт получаем 10 байт (80 бит) на пароль
5. на одного пользователя выделяется 900 : 30 = 30 байт
6. на дополнительную информацию остается 30 – 10 = 20 байт
7. ответ: 20.

### Ещё пример задания:

**Р-07.** *При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 15 символов и содержащий только символы Ш, К, О, Л, А (таким образом, используется 5 различных символов). Каждый такой пароль в компьютерной системе записывается минимально возможным и одинаковым целым количеством байт (при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально*

*возможным количеством бит). Укажите объём памяти в байтах, отводимый этой системой для записи 30 паролей. В ответе запишите только число, слово «байт» писать не нужно.*

**Решение**:

1. согласно условию, в пароле можно использовать 5 символов
2. для кодирования номера одного из 5 символов нужно выделить 3 бита памяти (они позволяют закодировать 23 = 8 вариантов)
3. для хранения всех 15 символов пароля нужно 15 ⋅ 3 = 45 бит
4. поскольку пароль должен занимать целое число байт, берем ближайшее большее (точнее, не меньшее) значение, которое кратно 8: это 48 = 6 ⋅ 8; то есть один пароль занимает 6 байт
5. тогда 30 паролей занимают 6 ⋅ 30 = 180 байт
6. ответ: 180.

### Ещё пример задания:

**Р-06.** *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 в байтах, который занимает хранение 60 паролей.*

**Решение**:

1. согласно условию, в пароле можно использовать 10 цифр (0..9) + 12 заглавных букв местного алфавита + 12 строчных букв, всего 10 + 12 + 12 = 34 символа
2. для кодирования номера одного из 34 символов нужно выделить 6 бит памяти (5 бит не хватает, они позволяют закодировать только 25 = 32 варианта)
3. для хранения всех 11 символов пароля нужно 11 ⋅ 6 = 66 бит
4. поскольку пароль должен занимать целое число байт, берем ближайшее большее (точнее, не меньшее) значение, которое кратно 8: это 72 = 9 ⋅ 8; то есть один пароль занимает 9 байт
5. тогда 60 паролей занимают 9 ⋅ 60 = 540 байт
6. ответ: 540.

|  |
| --- |
| **Возможные ловушки**:   * часто забывают, что пароль должен занимать ЦЕЛОЕ число байт |

### Ещё пример задания:

**Р-05.** *В велокроссе участвуют 119 спортсменов. Специальное устройство регистрирует прохождение каждым из участников промежуточного финиша, записывая его номер с использованием минимально возможного количества бит, одинакового для каждого спортсмена. Каков информационный объем в битах сообщения, записанного устройством, после того как промежуточный финиш прошли 70 велосипедистов?*

**Решение**:

1. велосипедистов было 119, у них 119 разных номеров, то есть, нам нужно закодировать 119 вариантов
2. по таблице степеней двойки находим, что для этого нужно минимум 7 бит (при этом можно закодировать 128 вариантов, то есть, еще есть запас); итак, 7 бит на один отсчет
3. когда 70 велосипедистов прошли промежуточный финиш, в память устройства записано 70 отсчетов
4. поэтому в сообщении 70\*7 = 490 бит информации.

|  |
| --- |
| **Возможные ловушки**:   * дано число, которое есть в условии (неверные ответы 70 бит, 70 байт, 119 байт), чтобы сбить случайное угадывание * указано правильное число, но другие единицы измерения (мог быть вариант 490 байт) * расчет на невнимательное чтение условия: можно не заметить, что требуется определить объем только 70 отсчетов, а не всех 119 (мог быть вариант 119\*7=833 бита) |

### Еще пример задания:

**Р-04.** *Объем сообщения, содержащего 4096 символов, равен 1/512 части Мбайта. Какова мощность алфавита, с помощью которого записано это сообщение?*

|  |
| --- |
| **Большие числа. Что делать?**  Обычно (хотя и не всегда) задачи, в условии которых даны большие числа, решаются достаточно просто, если выделить в этих числах степени двойки. На эту мысль должны сразу наталкивать такие числа как  128 = 27, 256 = 28, 512 = 29 , 1024 = 210,  2048 = 211, 4096 = 212 , 8192 = 213, 16384 = 214, 65536 = 216 и т.п.  Нужно помнить, что соотношение между единицами измерения количества информации также представляют собой степени двойки:  1 байт = 8 бит = 23 бит,  1 Кбайт = 1024 байта = 210 байта  = 210 · 23 бит = 213 бит,  1 Мбайт = 1024 Кбайта = 210 Кбайта  = 210 · 210 байта = 220 байта  = 220 · 23 бит = 223 бит.  Правила выполнения операций со степенями:   * при умножении степени при одинаковых основаниях складываются      * … а при делении – вычитаются: |

**Решение (вариант 1)**:

1. в сообщении было 4096 = 212 символов
2. объем сообщения

1/512 Мбайта = 223 / 512 бита = 223 / 29 бита = 214 бита (= 16384 бита!)

1. место, отведенное на 1 символ:

214 бита / 212 символов = 22 бита на символ = 4 бита на символ

1. 4 бита на символ позволяют закодировать 24 = 16 разных символов
2. поэтому мощность алфавита – 16 символов

|  |
| --- |
| **Возможные ловушки**:   * легко запутаться, если выполнять вычисления «в лоб», не через степени двойки |

**Решение (вариант 2, предложен В.Я. Лаздиным)**:

1. объем сообщения

1/512 Мбайт = 1024/512 Кбайт = 2 Кбайт = 2048 байт

1. на 1 символ приходится 2048 байт / 4096 = 1/2 байта = 4 бита
2. 4 бита на символ позволяют закодировать 24 = 16 разных символов
3. поэтому мощность алфавита – 16 символов

|  |
| --- |
| **Возможные проблемы**:   * не всегда удобно работать с дробными числами (1/2 байта) * метод разработан специально для этой задачи, где он хорошо работает; в других задачах может быть не так гладко |

### Еще пример задания:

**Р-03.** *В зоопарке 32 обезьяны живут в двух вольерах, А и Б. Одна из обезьян заболела. Сообщение «Заболевшая обезьяна живет в вольере А» содержит 4 бита информации. Сколько обезьян живут в вольере Б?*

**Решение (вариант 1)**:

1. информация в 4 бита соответствует выбору одного из 16 вариантов, …
2. … поэтому в вольере А живет 1/16 часть всех обезьян (это **самый важный момент**!)
3. всего обезьян – 32, поэтому в вольере А живет

32/16 = 2 обезьяны

1. поэтому в вольере Б живут все оставшиеся

32 – 2 = 30 обезьян

1. ответ – 30.

|  |
| --- |
| **Возможные ловушки**:   * можно сделать неверный вывод о том, что в вольере А живет 4 обезьяны (столько же, сколько бит информации мы получили), следовательно, в вольере Б живут оставшиеся 28 обезьян (неверный ответ 3) * после п. 1 можно сделать (неверный) вывод о том, что в вольере А живет 16 обезьян, следовательно, в вольере Б – тоже 16 (неверный ответ 2) |

**Решение (вариант 2, использование формулы Шеннона[[2]](#footnote-2))**:

1. заболевшая обезьяна может жить в вольере А (событие 1) или в вольере Б (событие 2)
2. количество информации в сообщении о произошедшем событии с номером  равно , где  – вероятность этого события; таким образом, получаем вероятность того, что заболевшая обезьяна живет в вольере А:

.

1. у нас не было никакой предварительной информации о том, где живет заболевшая обезьяна, поэтому можно считать, что вероятность определяется количеством обезьян в вольере – если вероятность равна 1/16, то в вольере живет 1/16 часть всех обезьян:

32/16 = 2 обезьяны

1. поэтому в вольере Б живут все оставшиеся

32 – 2 = 30 обезьян

1. ответ – 30.

### Еще пример задания:

**Р-02.** *В корзине лежат 32 клубка шерсти, из них 4 красных. Сколько бит информации несет сообщение о том, что достали клубок красной шерсти?*

**Решение (вариант 1)**:

1. красные клубки шерсти составляют 1/8 от всех, …
2. поэтому сообщение о том, что первый вынутый клубок шерсти – красный, соответствует выбору одного из 8 вариантов
3. выбор 1 из 8 вариантов – это информация в 3 бита (по таблице степеней двойки)
4. ответ – 3.

**Решение (вариант 2, использование формулы Шеннона)**:

1. красные клубки шерсти составляют 1/8 от всех, поэтому вероятность  того, что первый вынутый клубок шерсти – красный, равна 1/8
2. по формуле Шеннона находим количество информации в битах:

бита.

1. ответ – 3.

### Еще пример задания:

**Р-01.** *В некоторой стране автомобильный номер длиной 7 символов составляется из заглавных букв (всего используется 26 букв) и десятичных цифр в любом порядке. Каждый символ кодируется одинаковым и минимально возможным количеством бит, а каждый номер – одинаковым и минимально возможным целым количеством байт. Определите объем памяти, необходимый для хранения 20 автомобильных номеров.*

**Решение**:

1. всего используется 26 букв + 10 цифр = 36 символов
2. для кодирования 36 вариантов необходимо использовать 6 бит, так как , т.е. пяти бит не хватит (они позволяют кодировать только 32 варианта), а шести уже достаточно
3. таким образом, на каждый символ нужно 6 бит (минимально возможное количество бит)
4. полный номер содержит 7 символов, каждый по 6 бит, поэтому на номер требуется бита
5. по условию каждый номер кодируется целым числом байт (в каждом байте – 8 бит), поэтому требуется 6 байт на номер (), пяти байтов не хватает, а шесть – минимально возможное количество
6. на 20 номеров нужно выделить байт
7. ответ – 120.

|  |
| --- |
| **Возможные ловушки**:   * если не обратить внимание на то, что каждый номер кодируется целым числом БАЙТ, получаем неверный ответ 2 (бит = 105 байт) * если по невнимательности считать, что каждый СИМВОЛ кодируется целым числом байт, получаем 7 байт на символ и всего 140 байт (неверный ответ 4) * если «забыть» про цифры, получим всего 26 символов, 5 бит на символ, 35 бит (5 полных байт) на каждый номер и неверный ответ 100 байт (на 20 номеров) |

**Еще пример задания:**

**Р-00.** *В школьной базе данных хранятся записи, содержащие информацию об учениках:*

*<Фамилия> – 16 символов: русские буквы (первая прописная, остальные строчные),*

*<Имя> – 12 символов: русские буквы (первая прописная, остальные строчные),*

*<Отчество> – 16 символов: русские буквы (первая прописная, остальные строчные),*

*<Год рождения> – числа от 1992 до 2003.*

*Каждое поле записывается с использованием минимально возможного количества бит. Определите минимальное количество байт, необходимое для кодирования одной записи, если буквы е и ё считаются совпадающими.*

**Решение**:

1. очевидно, что нужно определить минимально возможные размеры в битах для каждого из четырех полей и сложить их;
2. важно! известно, что первые буквы имени, отчества и фамилии – всегда заглавные, поэтому можно хранить их в виде строчных и делать заглавными только при выводе на экран (но нас это уже не волнует)
3. таким образом, для символьных полей достаточно использовать алфавит из 32 символов (русские строчные буквы, «е» и «ё» совпадают, пробелы не нужны)
4. для кодирования каждого символа 32-символьного алфавита нужно 5 бит (32 = 25555), поэтому для хранения имени, отчества и фамилии нужно (16 + 12 + 16)•5=220 бит
5. для года рождения есть 12 вариантов, поэтому для него нужно отвести 4 бита (24 = 16 ≥ 12)
6. таким образом, всего требуется 224 бита или 28 байт
7. правильный ответ – 28 байт.

1. Часто килобайт обозначают «Кб», а мегабайт – «Мб», но в демо-тестах разработчики ЕГЭ привели именно такие обозначения. [↑](#footnote-ref-1)
2. Фактически это не другой способ решения, а более строгое обоснование предыдущего алгоритма. [↑](#footnote-ref-2)