Методические материалы для подготовки к ЕГЭ по информатике по теме «Информация и ее кодирование»

 Кодирование информации – одна из базовых тем курса информатики и ИКТ, отражающая фундаментальную необходимость представления информации в какой-либо форме.

Тема «Информация и её кодирование» очень значима в ЕГЭ по информатике. Проверке знаний и умений по разделу «Информация и её кодирование» содержания курса информатики посвящены задания: №1, № 16 (системы счисления), № 9, № 13 (оценка количественных характеристик информации), № 5 (однозначность двоичного кода) и № 10 (комбинаторика) из которых пять базового, и одно (№ 13) – повышенного уровня сложности.

**Базовые понятия:**

 **Информация** — это сведения об объектах и явлениях окружающей среды, их свойствах, уменьшающие неопределенность и/или неполноту знаний.

**Бит (Binary digIT)** — это единица измерения количества информации, равная количеству информации, содержащемуся в опыте, имеющем два равновероятных исхода.

**Кодирование информации —** это процесс однозначного преобразования информации с одного языка на другой. Однозначный процесс, значит имеющий правило/систему правил для обратного преобразования информации в первоначальный вид. Неоднозначный процесс, значит не позволяющий вернуться к первоначальному виду информации, искажающий ее.

**Декодирование информации —** это процесс преобразования информации обратный кодированию.

**Равномерное кодирование —** это кодирование, при котором все символы кодируются кодами равной длины.

**Неравномерное кодирование** — это кодирование, при котором разные символы могут кодироваться кодами разной длины.

**Алфавит —** это совокупность всех различных символов, которая используется для записи сообщения.

 **Глубина кодирования цвета —** это количество бит, необходимых для хранения и представления цвета при кодировании одного пикселя растровой графики.

**Базовые формулы:**

**N = 2i,** где **N** — это количество различных символов в алфавите, i — это минимальное количество информации (бит), которое требуется для кодирования одного символа в алфавите.

**I = K · i,** где **I —** это информационный объем сообщения в битах (байтах, Кбайтах…),

**K —** это количество символов в сообщении (для текстового сообщения К — это количество всех знаков в сообщении; для графического изображения: К — это количество пикселей в растровом изображении; для звукового файла: в формуле есть дополнительные множители, подробнее в других уроках),

**i —** это количество бит на кодирование одного символа (в терминологии кодирования графической информации i — глубина кодирования цвета).

**Необходимо четко знать:**

**- значения степеней числа 2;**

**- правила перевода из различных систем счисления в десятичную и обратно;**

**- единицы измерения информации**

1 байт = 23 бит

1 Кбайт = 210байт = 213бит

1Мбайт = 210Кбайт = 220байт = 223бит

1Гбайт = 210Мбайт = 220Кбайт = 230байт = 233бит

Поскольку на экзамене по информатике нельзя пользоваться калькулятором, то учимся вычислять выражения со степенями 2, не прибегая к сложным вычислениям с длинными числами.

**Задание. Вычислим, сколько бит содержится в** $\frac{1}{256}$**МБайт :**

**Решение.**

**1-ый способ:**

****

Любую арифметическую операцию умножения или деления всегда надо проверять. Здесь работаем с большими числами, высока вероятность ошибки.

**2-ой способ (более простой):**

****

Во втором способе решения мы только складываем и вычитаем значения степеней 2. Применяем основные формулы для преобразования степеней, которые будут полезны при решении многих заданий ЕГЭ.

При подготовке учащихся к ЕГЭ все задания по теме «Информация и кодирование» целесообразно разбить на 3 блока (на каждый блок отводим одно занятие):

I – системы счисления;

II – равномерное и неравномерное кодирование, правило Фано и комбинаторика;

III – расчет объема текстовой, графической и звуковой информации, скорости передачи данных.

1. Рассмотрим первый блок заданий ЕГЭ– позиционные системы счисления.

Считаю, что почти все задания ЕГЭ по информатике из года в год имеют тенденцию к усложнению. Это касается и раздела системы счисления. Сравним оба задания этого блока из демоверсий 2016 и 2017 г.

**№ 1.**

***2016:*** Сколько единиц в двоичной записи шестнадцатеричного числа E1A016?

***2017:*** Сколько существует натуральных чисел x, для которых выполнено неравенство 110111002 < x < DF16? В ответе укажите только количество чисел, сами числа писать не нужно.

Это задание наверно не случайно находится в КИМ-ах на первом месте, так как является наиболее простым. Для решения задания достаточно перевести число из шестнадцатиричной системы в двоичную, заменив каждую цифру ее двоичным 4-х битным представлением ([метод быстрого перевода чисел из двоичной системы в шестнадцатеричную](http://infoegehelp.ru/index.php?option=com_content&view=article&id=56:perevod-chisel-bin-oct-hex&catid=39&Itemid=64)). В 2017 году проверяется также умение сравнивать числа в 2-ичной или 16-тиричной системах.

Более сложным является задание 16.

**№ 16**

***2016:*** Сколько единиц содержится в двоичной записи значения выражения: 42013 + 22012 – 16?

***2017:*** Значение арифметического выражения: 918 + 354 – 9 – записали в системе счисления с основанием 3. Сколько цифр «2» содержится в этой записи?

Рассмотрим на примерах, как решаются подобные задачи.

***Пример1.***



***Пример 2.***

Десятичное число 57 в некоторой системе счисления записывается как 212. Определите основание системы счисления

 Решение:

 5710=212х

57=2\*х2+1\*х1+2\*х0

Решая квадратное уравнение 2х2+х-55 =0 получаем х=5.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Пример 3.***



Для решения используем правила перевода в различные системы счисления чисел, равных основанию системы в n-ой степени:



Чтобы посчитать сколько единиц содержится в двоичной записи числа х, его можно представить в виде разности:



Вернемся к решению примера 3:



Ответ: 4400.

***Пример 4.***



***Пример 5.***



1. Следующий тип задач – это расшифровка сообщений, закодированных с помощью неравномерного двоичного кода, удовлетворяющего условию Фано. К такому типу относится задание **№5.**

***2016:*** По каналу связи передаются сообщения, содержащие только четыре буквы: А, Б, В, Г; для передачи используется двоичный код, удовлетворяющий условию Фано. Для букв А, Б, В используются такие кодовые слова: А – 0; Б – 110; В – 100.

Укажите кратчайшее кодовое слово для буквы Г, при котором код будет допускать однозначное декодирование. Если таких кодов несколько, укажите код с наименьшим числовым значением.

***2017:*** Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г, Д, Е, решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для буквы А использовали кодовое слово 0; для буквы Б – кодовое слово 10. Какова наименьшая возможная сумма длин всех шести кодовых слов?

**Примечание.** Условие Фано означает, что никакое кодовое слово не является началом другого кодового слова. Это обеспечивает возможность однозначной расшифровки закодированных сообщений.

 Решение данного типа задач не требует глубоких знаний и формул, а сводится к построению «битового» дерева, наглядно демонстрирующего различные двоичные цепочки для кодировки букв из сообщения.

Рассмотрим это на примерах.

***Пример 1.*** (К. Поляков)

 

**Решение:**

В соответствии с условием Фано, код буквы С не может начинаться на 0. На построенной схеме видно, что существует два варианта ответа: 101 и 110. Исходя из условия задачи выбираем код с наименьшим числовым значением: 101.

***Пример 2.*** (К. Поляков)



**Решение.**

По условию Фано и на схеме показано, что искомые коды согласных букв не могут начинаться на 1, на 01 и на 001.

Они могут начинаться только с 000. Трехбитный код 000 возможен только для одной буквы, четырехбитный – для двух вариантов букв (0000 и 0001), пятибитный –для четырех букв, а по условию их пять. Значит ответ – 6 бит.

***Пример 3.***

Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г, решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для буквы А использовали кодовое слово 0, для буквы Б – кодовое слово 110.

Какова наименьшая возможная суммарная длина всех четырёх кодовых слов?

**Решение.**

Построим дерево для заданных кодовых слов А – 0 и Б – 110:

Штриховыми линиями отмечены две «пустые» ветви, на которые можно «прикрепить» листья для кодовых слов букв В (10) и Г (111). Получаем суммарную длину кодовых слов

9 символов.

***Пример 4.***

Решим задание № 5 из демоверсии 2016г.

**Решение.**

Построив дерево, получаем два варианта ответа, из которых выбираем наименьшее значение кода для буквы Г. Это - **101.**

***Пример 5.***

Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г и Д, используется неравномерный двоичный код, позволяющий однозначно декодировать полученную двоичную последовательность. Вот этот код:

А – 0; Б – 100; В – 1010; Г – 111; Д – 110. Требуется сократить для одной из букв длину кодового слова так, чтобы код по-прежнему можно было декодировать однозначно. Коды остальных букв меняться не должны.

**Решение.**

Изобразим на дереве цепочки с кодами букв последовательности из условия задачи:

А – 0; Б – 100; В – 1010; Г – 111; Д – 110.

По полученному рисунку делаем вывод, что код буквы В может быть сокращен с четырехбитного варианта, до трехбитового, а именно вместо 1010 можно использовать цепочку битов **101.**

Разобрав данные примеры учащиеся уже самостоятельно легко справляются с решением аналогичных задач из открытого банка заданий различных образовательных ресурсов.

**Задание № 10** предполагает проверку знаний теоретического материала и умений решать задачи из комбинаторики.

**Теория:**



Рассмотрим несколько примеров.

***Пример 1.***Сколь­ко слов длины 5 можно со­ста­вить из букв Е, Г, Э? Каж­дая буква может вхо­дить в слово не­сколь­ко раз.

**Решение:**

Если в алфавите *M* символов, то количество всех возможных «слов» (сообщений) длиной *N* равно *Q* = *MN*.

 В нашем слу­чае *N* = 5, *M* = 3. Сле­до­ва­тель­но, *Q* = 35 = 243.

**Ответ**: 243

 ***Пример 2.***

Вася составляет 5-буквенные слова, в которых есть только буквы С, Л, О, Н, причём буква С используется в каждом слове ровно 1 раз. Каждая из других допустимых букв может встречаться в слове любое количество раз или не встречаться совсем.

 Словом считается любая допустимая последовательность букв, не обязательно осмысленная. Сколько существует таких слов, которые может написать Вася?

**Решение:**

Буква С может стоять на одном из пяти мест: С\*\*\*\*, \*С\*\*\*, \*\*С\*\*, \*\*\*С\* и \*\*\*\*С, где \* обозначает любой из оставшихся трёх символов в каждом случае в остальных четырёх позициях может быть любая из трёх букв Л, О, Н, поэтому при заданном расположении буквы С имеем 34 = 81 вариант всего вариантов 5 · 81 = 405.

**Ответ:** 405.

***Пример 3.***

Сколько существует различных символьных последовательностей длины 5 в четырёхбуквенном алфавите {A, C, G, T}, которые содержат ровно две буквы A?

**Решение:**

Рассмотрим различные варианты слов из 5 букв, которые содержат две буквы А и начинаются с А: АА\*\*\* А\*А\*\* А\*\*А\* А\*\*\*А

Звёздочка обозначает любой символ из набора {C, G, T}, то есть один из трёх символов. Итак, в каждом шаблоне есть 3 позиции, каждую из которых можно заполнить тремя способами, поэтому общее число комбинаций (для каждого шаблона!) равно 33 = 27

 Всего 4 шаблона, они дают 4 · 27 = 108 комбинаций.

Теперь рассматриваем шаблоны, где первая буква А стоит на второй позиции, их всего три: \*АА\*\* \*А\*А\* \*А\*\*А. Они дают 3 · 27 = 81 комбинацию.

Два шаблона, где первая буква А стоит на третьей позиции: \*\*АА\* \*\*А\*А. Они дают 2 · 27 = 54 комбинации и один шаблон, где сочетание АА стоит в конце \*\*\*АА они дают 27 комбинаций. Всего получаем (4 + 3 + 2 + 1) · 27 = 270 комбинаций **.**

**Ответ:** 270**.**

***Пример 4.***

Все 4-буквенные слова, составленные из букв К, Л, Р, Т, записаны в алфавитном порядке и пронумерованы. Вот начало списка:

1. КККК

2. КККЛ

3. КККР

4. КККТ

……

Запишите слово, которое стоит на 67-м месте от начала списка.

**Решение:**

Самый простой вариант решения этой задачи – использование систем счисления; Действительно, здесь расстановка слов в алфавитном порядке равносильна расстановке по возрастанию чисел, записанных в четверичной системе счисления (основание системы счисления равно количеству используемых букв).

Выполним замену К→0, Л→1, Р→2, Т→3; поскольку нумерация слов начинается с единицы, а первое число КККК→0000 равно 0, под номером 67 будет стоять число 66, которое нужно перевести в четверичную систему: 66 = 10024

Выполнив обратную замену (цифр на буквы), получаем слово ЛККР.

**Ответ:** ЛККР.

1. Последний блок заданий посвящен проверке умений производить количественную оценку информации различного типа.

 Задание № 9 из КИМов ЕГЭ проверяет знание способов кодирования растровых изображений и алгоритма нахождения объема графического файла. Решение такого вида заданий не вызывают затруднений у выпускников и достаточно хорошо отработаны ими в процессе обучения в курсе основной школы.

№ 9

***2016:*** Какой минимальный объём памяти (в Кбайт) нужно зарезервировать, чтобы можно было сохранить любое растровое изображение размером 320×640 пикселей при условии, что в изображении могут использоваться 256 различных цветов? В ответе запишите только целое число, единицу измерения писать не нужно.

**Решение**.

По условию К=320\*640 пикселей, N=256, значит глубина цвета по формуле Хартли i=8 (т.к. 28=256). Тогда объем файла I=K\*I, I = 320\*640\*8 бит =25\*10\*26\*10\*23 бит. Переводим в Кбайты и получаем: 100\*214/213=50 Кбайт.

**Ответ:** 50 Кбайт.

***2017:***  Для хранения произвольного растрового изображения размером 1024×1024 пикселей отведено 512 Кбайт памяти, при этом для каждого пикселя хранится двоичное число – код цвета этого пикселя. Для каждого пикселя для хранения кода выделено одинаковое количество бит. Сжатие данных не производится. Какое максимальное количество цветов можно использовать в изображении?

**Решение.**

По условию К= 210\*210 пикселей, I=512 Кбайт или 29\*213 бит. Тогда глубина цвета i=I/K=222/220=22=4 бит. Следовательно, по формуле Хартли получаем N=24=16 цветов.

**Ответ:** 16 цветов.

 В задания КИМов ЕГЭ последних лет не включены задания на определение объема звукового файла и скорости передачи данных по сети. Вероятно, из-за простоты их решения, эти задачи содержатся в ОГЭ для выпускников 9 класса.

 Задание № 13 считается заданием повышенного уровня сложности. Решается с использованием алфавитного подхода к измерению количества информации.



Данные задачи также неплохо представлены в школьном курсе обучения информатике, но в более простом варианте, как например в задаче из демоверсий прошлых лет:

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 занимает хранение 60 паролей.

**Решение.**

1**.** Определяется мощность используемого алфавита: 10 цифр и 12 букв, каждая из которых может иметь два возможных начертания. Итого 10+12\*2=34 знака.

2. Определяется **целое** количество битов, минимально достаточное для представления одного знака такого алфавита. Поэтому выбирается ближайшее большее число N, равное степени числа 2. В нашем случае это 26=64 (значения 25=32 недостаточно). Значит i=6 бит на один знак алфавита.

3. Длина пароля равна11 символам (К=11). Тогда объем одного пароля I=К\*i=11\*6=66 битов.

4. Под хранение каждого пароля отводится минимальное целое количество байтов. Значит 66:8=8,259 байтов.

5. Тогда под хранение 60 паролей потребуется 60\*9=540 байтов.

**Ответ**: 540 байтов.

 В последние годы задание № 13 усложнилось добавлением для каждого пользователя дополнительных сведений, объем которых и нужно найти.

***2016:*** При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 9 символов и содержащий только символы из 26-символьного набора прописных латинских букв. 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 20 пользователях потребовалось 400 байт. Сколько байт выделено для хранения дополнительных сведений об одном пользователе? В ответе запишите только целое число – количество байт.

**Решение.**

После нахождения мощности алфавита (N=9+26=35 знака), целого количества бит для каждого символа (i=6 бит, т.к.26=64) и длины каждого пароля в байтах (9\*6=54 бита или 7 байт), рассчитываем **целое** количество байт, отводимое для хранения пароля и дополнительных сведений о каждом пользователе: I1=400/20=20 байт. Оставшиеся 20-7=13 байт отводятся под хранение дополнительных сведений о каждом пользователе.

**Ответ:** 13 байт.

***2017:***  При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 9 символов. Из соображений информационной безопасности каждый пароль должен содержать хотя бы 1 десятичную цифру, как прописные, так и строчные латинские буквы, а также не менее 1 символа 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 сведения, для чего выделено целое число байт; это число одно и то же для всех пользователей. Для хранения сведений о 20 пользователях потребовалось 500 байт. Сколько байт выделено для хранения дополнительных сведений об одном пользователе? В ответе запишите только целое число – количество байт.

**Примечание.** В латинском алфавите 26 букв.

(**Ответ:** 17 байт).

Банк заданий для данного блока довольно обширен (см. список ЭОР)

 Умение решать задачи по информатике и ИКТ является важным критерием сформированности мыслительных способностей учащихся. Раздел «Информация и её кодирование», при правильной организации учебного процесса позволяет сформировать

у учащихся глубокие навыки решения заданий ЕГЭ по информатике.

**Используемые материалы и ЭОР:**

1. Методические рекомендации на сайте Полякова К. Ю. <http://kpolyakov.spb.ru> (ПОЛЯКОВ Константин Юрьевич, д.т.н., учитель информатики ГБОУ СОШ № 163, г. Санкт-Петербург, kpolyakov@mail.ru).
2. Лещинер В.Р. ЕГЭ 2015. Информатика. Типовые тестовые задания. — М.: Экзамен, 2015.
3. Крылов С.С., Чуркина Т.Е. ЕГЭ 2015. Информатика и ИКТ. Типовые экзаменационные варианты. — М.: «Национальное образование», 2015.
4. [www.fipi.ru](http://www.fipi.ru) – официалный сайт Федерльного института педгогических исследований
5. <http://ege-go.ru> М.А. Ройтберг mroytberg@lpm.org.ru
6. <http://infbu.ru/catalog/1013> - БУ информатик
7. <http://pandia.ru/text/78/122/98265.php>
8. <http://down.ctege.info/ege/obshee/inform/praktika/inform-a1-praktika.pdf>
9. <http://easyinformatics.ru/>
10. <http://infoegehelp.ru/>