

Тема 2

Кодирование данных

Тема 2, 3 из учебного
пособия
«Информационное и
компьютерное
обеспечение»

Кодирование данных двоичным

КОДОМ

Двоичное кодирование основано на представлении данных последовательностью двух знаков: **0** и **1**. Эти знаки называются *двоичными цифрами*, по – английски – *binary digit*, или, сокращенно, *bit* (*бит*).

Бит – двоичный разряд, принимающий значение 0 или 1.

1 бит – 2 значения: 0, 1

2 бит – 4 значения: 00, 01, 10, 11

3 бит – 8 значений: 000, 001, 010, 011, 100, 101, 110, 111

N бит – 2^N

Кодирование целых и действительных чисел

Для кодирования целых чисел от 0 до 255 достаточно иметь 8 разрядов (8 бит).

$$0000\ 0000 = 0$$

$$0000\ 0001 = 1$$

.....

$$1111\ 1110 = 254$$

$$1111\ 1111 = 255$$

16 бит позволяют закодировать целые числа от 0 до 65535, а 24 бита – уже более 16,5 миллионов разных значений.

Для кодирования **действительных чисел** число предварительно преобразуется в **форму с плавающей точкой**:

<Мантисса> E <Порядок>

Привычная для человека запись действительных чисел в виде

<Целая часть> . <Дробная часть>

называется **форма с фиксированной точкой**, например 12.36.

Запись этого числа **в форме с плавающей точкой**: 1.236E+01, или 0.1236E+02, или 123.6E-01.

Внутреннее представление числа в компьютере – в форме с плавающей точкой.

Действительное число может быть ***нормализованным***. Для этого:

- мантисса числа должна быть меньше 1;
- Первая цифра после запятой отлична от нуля.

Например, форма числа

$0.25E+02$ – нормализованная, а

$2.5E-01$ или $0,025E+03$ –
ненормализованные.

Единицы представления данных

Наименьшей единицей представления является **бит** (*двоичный разряд*). Группы из восьми битов называются **байтами**. Байт – минимальная адресуемая ячейка памяти.

Группа из 16 взаимосвязанных битов (двух взаимосвязанных байтов) в информатике называется **словом**. Соответственно, группы из четырех взаимосвязанных байтов (32 бита) называются **удвоенным словом**.

Единицы измерения данных

Наименьшей единицей измерения является **байт** (минимальный объем памяти со своим адресом).

Более крупные единицы измерения данных образуются добавлением префиксов **кило-**, **мега-**, **гига-**, **тера-**, **пета-**.

1 Кбайт = 1024 байт = 2^{10} байт

1 Мбайт = 1024 Кбайт = 2^{20} байт

1 Гбайт = 1024 Мбайт = 2^{30} байт

1 Тбайт = 1024 Гбайт = 2^{40} байт

1 Пбайт = 1024 Тбайт = 2^{50} байт

Единицы хранения данных

В качестве единицы хранения данных принят объект называемый *файлом*. **Файл** – это последовательность произвольного числа байтов, обладающая уникальным собственным именем.

Поскольку в определении файла нет ограничений на размер, можно представить себе файл, имеющий **0 байтов** (*пустой файл*), и файл, имеющий любое число байтов.

Понятие о файловой структуре

Хранение файлов организуется в иерархической структуре, которая называется *файловой структурой*.

В качестве **вершины** структуры служит имя носителя, на котором сохраняются файлы. Далее файлы группируются в *каталоги (папки)*, внутри которых могут быть созданы *вложенные каталоги (папки)*.

Путь доступа к файлу начинается с имени устройства и включает все имена каталогов (папок), через которые он проходит. В качестве разделителя используется символ «\» (обратная косая черта).

Уникальность имени файла обеспечивается тем, что *полным именем файла* считается собственное имя файла вместе с путем доступа к нему:

**<Путь доступа к файлу> \ <Имя .
расширение>**

Пример записи полного имени файла:

Пример записи двух файлов, имеющих одинаковое собственное имя и размещенных на одном носителе, но отличающихся путем доступа, то есть полным именем:

C:\Мои документы\Лекции\Физика\Лекция
1.doc

Кодирование текстовых данных

Для кодирования текстов используются различные таблицы перекодировки.

Таблица перекодировки – таблица, содержащая упорядоченный перечень кодируемых символов, в соответствии с которой происходит преобразование символа в его двоичный код и обратно.

Институт стандартизации США ввел в действие **систему кодирования *ASCII***.

В системе ASCII закреплены две таблицы кодирования: базовая и расширенная. Базовая таблица закрепляет значения кодов от 0 до 127, а расширенная – от 128 до 255.

Первые 32 кода базовой таблицы, начиная с нулевого, отданы производителям аппаратных средств (управляющие коды).

Начиная с кода 32 по код 127 размещены коды символов английского алфавита, знаков препинания, цифр, арифметических действий и некоторых вспомогательных символов.

В расширенной таблице, определяющую значения кодов со 128 по 255, размещены коды символов национального алфавита.

Кодировка символов русского языка, известная как кодировка **Windows– 1251**, была введена «извне» – компанией Microsoft.

Другая распространенная кодировка носит название **КОИ – 8** (код обмена информацией, восьмизначный).

Система, основанная на **16– разрядном** кодировании символов, получила название *универсальной* – **UNICODE**. Шестнадцать разрядов позволяют обеспечить уникальные коды для 65536 различных символов – этого поля достаточно для размещения в одной таблице символов большинства языков планеты.

Кодирование графических данных

При кодировании изображения происходит его пространственная **дискретизация**, т.е. разбиения непрерывного графического изображения на отдельные элементы.

Все изображение разбивается на отдельные точки, каждому элементу ставится в соответствие код его цвета.

Качество кодирования будет зависеть от следующих параметров: **размера точки** и **количества используемых цветов**.

Чем меньше размер точки, а, значит, изображение составляется из большего количества точек, тем выше качество кодирования.

Чем большее количество цветов используется, тем больше информации несет каждая точка, а, значит, увеличивается качество кодирования.

Создавать и хранить графические объекты в компьютере можно двумя способами – как **растровое** и **векторное** изображение. Отдельным предметом считается **3D (трехмерная) графика**, в которой сочетаются векторный и растровый способы формирования изображений.

Для каждого вида используется свой способ кодирования графической информации.

Растровое изображение

Графическое изображение состоит из мельчайших точек (*пикселей*), образующих характерный узор, называемый *растром*.

Черно– белые иллюстрации представляются в виде комбинации точек с **256** градациями серого цвета, и, т.о., для кодирования яркости любой точки (индивидуальные свойства точки) обычно достаточно **восемьразрядного** двоичного числа.

Для кодирования **цветных** графических изображений используется *принцип декомпозиции* произвольного цвета на основные составляющие.

Применяют несколько систем кодирования: **HSB**, **RGB** и **CMYK**.

Первая цветовая модель удобна для человека, вторая наиболее удобна для компьютера, а последняя модель CMYK – для типографий.

1) Модель ***HSB*** характеризуется тремя компонентами: **оттенок** цвета (Hue), **насыщенность** цвета (Saturation) и **яркость** цвета (Brightness).

Эту цветовую модель лучше применять в тех графических редакторах, в которых изображения создают сами, а не обрабатывают уже готовые.

Затем созданное свое произведение можно преобразовать в цветовую модель RGB, если ее планируется использовать в качестве экранной иллюстрации, или CMYK, если в качестве печатной.

2) Принцип модели **RGB** заключается в следующем: любой цвет можно представить в виде комбинации трех цветов: **красного** (Red, R), **зеленого** (Green, G), **синего** (Blue, B).

Другие цвета и их оттенки получаются за счет наличия или отсутствия этих составляющих.

По первым буквам основных цветов система и получила свое название – RGB.

Любой цвет можно получить сочетанием основных цветов в различных

Модель **СМУК** используется при подготовке публикаций к печати.

Каждому из основных цветов ставится в соответствие дополнительный цвет (дополняющий основной до белого).

Дополнительными цветами для красного является голубой (Cyan, C), для зеленого – пурпурный (Magenta, M), для синего – желтый (Yellow, Y). Используют еще черный цвет (Black, так как буква B уже занята синим цветом, то обозначают буквой K). Это связано с тем, что наложение друг на друга дополнительных цветов не дает чистого черного цвета.

Различают несколько **режимов** представления цветной графики:

- полноцветный (True Color);
- High Color;
- индексный.

При **полноцветном режиме** для кодирования яркости каждой из составляющих используют по 256 значений (8 двоичных разрядов), то есть на кодирование цвета одного пикселя (в системе RGB) надо затратить $8 * 3 = 24$ разряда. Это позволяет однозначно определять **16,5 МЛН ЦВЕТОВ**.

При кодировании с помощью системы CMYK для представления цветной графики надо иметь $8 * 4 = 32$ двоичных разряда.

Режим High Color – это кодирование при помощи **16– разрядных** двоичных чисел. Это позволяет однозначно определять **65536 цветов**.

При *индексном кодировании* цвета можно передать всего лишь 256 цветовых оттенков. Но так как 256 значений не передают весь диапазон цветов, то подразумевается, что к графическим данным прилагается палитра (справочная таблица). Сам код точки растра в данном случае означает не сам по себе цвет, а только его номер (индекс) в палитре.

В видеопамяти любое изображение представляется в **растровом виде**. Экран монитора разбит на фиксированное число пикселей.

Разрешающая способность экрана монитора – размер раstra, задаваемого в виде произведения $M * N$, где M – количество точек по горизонтали, N – количество точек по вертикали (число строк).

Количество цветов, воспроизводимых на экране монитора (K), и число бит, отводимых в видеопамяти под каждый пиксель (b) (битовая глубина), связаны формулой: $K=2^b$.

Объем видеопамяти, необходимой для хранения

Векторное изображение

Векторное изображение – это графический объект, состоящий из элементарных отрезков и дуг. Базовым элементом изображения является *линия*. Все прочие объекты векторной графики составляются из линий.

Так как линия описывается математически как единый объект, то и объем данных для отображения объекта средствами векторной графики значительно меньше, чем в растровой графике. Информация о векторном изображении кодируется как обычная буквенно – цифровая.

Контрольные вопросы

1. Что такое двоичное кодирование, что называется битом?
2. Сколько различных символов можно закодировать n битами, одним байтом?
3. Сколько двоичных разрядов требуется для кодирования 256, 65536 целых чисел?
4. Что представляет собой форма с фиксированной точкой, форма с плавающей точкой и нормализованная форма для действительных чисел?
5. Перечислите единицы представления данных, чему они равны.
6. Перечислите единицы измерения данных, чему они равны.

7. Что является единицей хранения данных?
Определение.
8. Что такое файловая структура?
9. Что является вершиной файловой структуры?
10. Что представляет собой путь доступа к файлу?
11. Что понимается под полным именем файла?
12. Какие системы кодирования текстовой информации вы знаете?
13. Что вы знаете о системе кодирования ASCII?
14. Какие типы изображений вы знаете? Их отличия.
15. Системы кодирования графических данных
16. Какие существуют режимы представления цветной графики?