

Измерение и кодирование информации

Методические материалы, статьи

Обзорный материал для подготовки учащихся к рубежному тестированию по информатике по темам "Единицы измерения количества информации", "Определение количества информации" и "Кодирование информации".

Измерение и кодирование информации

Единицы информации

Бит - минимальная единица информации, представляющая собой наименьшую "порцию" памяти - 1 двоичный разряд. Бит обозначает количество информации, необходимое для различения двух равновероятных событий - значение размером в 1 бит представляет собой сообщение, уменьшающее неопределенность знания в два раза.

Байт - основная единица информации.

1 байт = 8 бит; 1 Кбайт = 2^{10} байт = 1024 байт; 1 Мбайт = 2^{10} Кбайт = 1024 Кбайт; 1 Гбайт = 2^{10} Мбайт = 1024 Мбайт

Определение количества (объема) информации

Вероятностный подход к измерению количества информации - рассматривает информацию с точки зрения повышения определенности знания в результате ее получения:

Количество (объем) получаемой информации (I) при известном количестве возможных событий (N):

Для равновероятных событий: $I = \log_2 N$

Для разновероятных событий: (формула Шеннона),

I - кол-во информации (бит), N - кол-во возможных событий, P_i - вероятность i-го события, $i = 1, 2, \dots, N$.

Алфавитный подход к измерению количества информации - рассматривает информацию с точки зрения ее представления как некоторой последовательности "знаков" (букв, цифр, кодов цвета точек изображения и т. п.). Алфавит знаковой системы - полный набор "знаков", используемых для формирования сообщений в данной знаковой системе. Объем информации, который несет каждый "знак" (информационная емкость "знака", объем занимаемой памяти): $I = \log_2 N$, где N - кол-во знаков в алфавите.

Количество (объем) информации, которое содержит сообщение, закодированное с помощью знаковой системы, равно: Информационная емкость "знака" * Количество знаков в сообщении.

Представление информации в компьютере

Все виды информации в компьютере обрабатываются в двоично-кодированном виде - т. е. в виде последовательности нулей и единиц, на физическом уровне представляемой в форме электрических импульсов: 1 - есть импульс, 0 - нет импульса.

Логические последовательности нулей и единиц представляют собой машинный язык.

Кодирование текстовой информации

Принцип кодирования: каждому символу ставится в соответствие определенный уникальный числовой (двоичный) код. Таблица, устанавливающая такое соответствие, называется таблицей кодировки символов.

Количество различных символов (N), которые можно закодировать с помощью какой-либо таблицы кодировки, определяется числом двоичных разрядов (k), отводимых под кодирование одного символа: $N=2^k$. Наибольшее распространение получило 8-разрядное кодирование (на кодирование одного символа отводится 8 бит = 1 байт), позволяющее закодировать $N=2^8=256$ различных символов.

В различных кодировках: коды от 0 до 32 - соответствуют управляющим клавишам (Enter, Esc и т. п.); от 33 до 127 - латинским буквам, цифрам, знакам арифметическим операций и знакам препинания; от 128 до 255 (расширенная часть таблицы кодировки) - для символов национальных алфавитов (в т. ч. кириллицы).

Наиболее распространенные 8-разрядные таблицы кодировок: ASCII (принята в качестве стандарта в MS-DOS), Windows-1251 (CP1251), КОИ-8, ISO.

UNICODE - 16-разрядная кодировка символов, позволяющая закодировать $2^{16} = 65536$ различных символов.

Кодирование графической информации

Минимальный объект кодирования растрового графического изображения - пиксель.

В основе кодирования цветных графических изображений - принцип декомпозиции цветов - т. е. разложение произвольного цвета на основные составляющие (например, по системе RGB: красный (Red), зеленый (Green) и синий (Blue)).

Глубина кодирования (глубина цвета) - количество бит (двоичных разрядов), используемых для кодирования цвета одной точки. От глубины цвета (k) зависит количество отображаемых цветов (N) - т. е. количество возможных состояний одной точки изображения: $N=2^k$.

Наиболее распространенные значения глубины цвета: 4, 8, 16, 24 бита на точку.

Разрешение - количество точек (пикселей) изображения, приходящихся на единицу длины. От разрешения зависит размер пикселя.

Наиболее частот используемые экранные разрешения: 640x480, 800x600, 1024x768, 1280x1024 точек.

Глубина кодирования и разрешение влияют на качество кодирования изображения.

Объем видеопамати (V), необходимый для формирования графического изображения на экране: $V = M * N * k$, где M - кол-во точек изображения по горизонтали, N - кол-во точек изображения по вертикали, k - глубина цвета (бит).

Кодирование звуковой информации

Для кодирования непрерывного (аналогового) звукового сигнала производится его дискретизация по времени (временная дискретизация, оцифровка) - разбиение непрерывной звуковой волны на отдельные короткие временные участки с измерением для каждого из них интенсивности звукового сигнала (величины амплитуды). Это выполняется аналогово-цифровым преобразователем (АЦП). При воспроизведении закодированного (оцифрованного) звука выполняется обратное преобразование цифро-аналоговым преобразователем (ЦАП) с последующим сглаживанием ступенчатого сигнала через аналоговый фильтр.

Глубина кодирования звука - количество бит (двоичных разрядов), используемых для кодирования уровня интенсивности (амплитуды) одного звукового сигнала. От глубины звука (k) зависит количество отражаемых в кодировании различных уровней звукового сигнала (N): $N=2^k$.

Частота дискретизации - количество измерений уровня (амплитуды, интенсивности) звукового сигнала в единицу времени. Количество измерений в секунду может лежать в диапазоне от 8000 до 48000, т. е.

диапазон частоты дискретизации аналогового звукового сигнала: от 8 до 48 кГц.

Глубина кодирования и частота дискретизации влияют на качество кодирования звука.

Информационный объем моноаудиофайла (V): $V = N * f * k$, где N - общая длительность звучания (секунд), f - частота дискретизации (Гц), k - глубина кодирования (бит).

Представление чисел в памяти компьютера

Традиционная форма записи чисел, используемая в повседневной жизни, называется естественной формой записи чисел. Существует также экспоненциальная форма представления чисел: $A_q = m * q^p$, где q - основание системы счисления, m - мантисса числа, p - порядок числа. Для 10-ой с/с: $A_{10} = m * 10^p$, для 2-ой с/с: $A_2 = m * 2^p$.

В компьютере числа представляются в одной из двух форм:

1. В форме с фиксированной точкой - соответствует естественной двоичной форме записи чисел с фиксированной разрядностью и указанием знака числа. В современных ЭВМ в такой форме представляются только целые числа.
2. В форме с плавающей точкой - соответствует экспоненциальной двоичной форме записи чисел с фиксированной разрядностью мантиссы и порядка и указанием знаков мантиссы и порядка. В компьютере числа в плавающей форме записываются в нормализованном виде (когда первая цифра мантиссы числа не равна нулю).

См. также:

[Курсовая работа на заказ по психологии детей дошкольного возраста](#)
[Помощь в учебе через Интернет](#)