

Информатика

Шмаков И. А.

7 октября 2019 г.

№18

Лекция 1

Информатика – это наука изучающая информационные аспекты процессов и системные аспекты информационных процессов.

Термин впервые появился в 1957 году благодаря Карлу Штейнбуху. В 1962 Дрейфусом во Франции. Харкувич в 1962 в СССР.

Информация – сведения или объект о чем-то **Объем данных** – кол-во байт, необходимых для их хранения в памяти электронного носителя. Бит – базовая единица измерения кол-ва информации.

Машинное слово – машино-зависящее и платформо-зависящее величина, измеряющаяся в битах или байтах.

Перевод из одной системы счисления в другую:

$10_{10} \rightarrow N_2$, делим число на 2 и ее остаток пока не получим 1 и дальше делить не можем и записываем в обратном порядке остатки.

Двоичное представление:

10:ABCD

0:0000

1:0001

2:0010

...

8:1000

9:1001

10:1010

11:1011

12:1100

13:1101

14:1110

15:1111

Для возвращения в 10ную систему счисления нужно возвести в степень $(1001111_2 = 1 \cdot 2^6 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 79_{10})$

Лекция 2 – Информация, кодирование информации, код Шенона и различные кодировки

Данные – подающееся многократной интерпретации, представление информации в формализованном виде, пригодном для передачи, связи или обработки.

Свойства:

1. Объективность
2. Достоверность
3. Полнота – минимальный набор, достаточный для принятия решений
4. Адекватность
5. Доступность
6. Актуальность (только вовремя полученная информация является полезной)
7. Ценность
8. Понятность (ясность)
9. Точность
10. Атрибутивные св.
11. Динамические св.
12. Практические св.

Теория информации – раздел прикладной математики, относящийся к измерению кол-ва информации, ее свойств и устанавливающий предельные соотношения для систем передачи данных.

Схема передачи информации

Источник информации → кодер источника → кодер канала → модулятор → среда распространения → демодулятор → декодер канала → декодер источника → получатель информации

Передача информации – это заблаговременно организованное техническое мероприятие, результатом которого становится воспроизведение

информации, имеющейся в одном месте, в другое место.

Информационная энтропия – мера неопределенности или непредсказуемости некоторой системы, в частности неопределенность появления какого-либо символа первичного алфавита.

Энтропия – это количество информации, приходящейся на одно элементарное сообщение источника, вырабатывающего ...

Формула Хартли:

$$l = \log_2(N) = n \log_2(m)$$

N - кол-во возможной информации

m - кол-во букв в алфавите

n - кол-во букв в сообщении

l - кол-во информации в битах

это верно при равновероятном появлении символа

Информационная двоичная энтропия для независимых случайных событий:

$$H(x) = -K \sum_{i=1}^n p(i) \log_2 p(i)$$

Лекция 3. Кодирование данных

Прямой код – способ представления двоичных чисел с фиксированной

Сумматор – устройство преобразующее информационные сигналы в сигнал, эквивалентный сумме этих сигналов, устройство производящее операцию сложения.

1938 году в "Bell laboratories" создали первый электронный двоичный сумматор.

Двоичный сумматор может быть описан с помощью:

- таблицы истинности
- в виде формулы

– в виде логической схемы

Обратный код – метод вычислительной математики, позволяющий вычесть одно число из другого используя только операцию сложения над натуральными числами. Обратный код положительного числа совпадает с Прямым кодом.

Дополнительный код – он позволяет заменить операцию вычитания на операцию сложения и сделать операции сложения и вычитания одинаковыми для знаковых и беззнаковых чисел.

Форма представления чисел с плавающей точкой состоит с помощью мантисы и показателя степени.

Машинный эпсилон – наименьшее положительное число, такое что неравное 1.

Кодирование графических данных:

- 1) Растровое – сетка из пикселей
- 2) Векторное – представление объектов с помощью примитивов (точки, линии, спрайты и т.д.)
- 3) Фрактальная – состоит из фракталов, объектов, отдельные элементы которого наследуют свойства родительской структуры.

Лекция 4. Кодирование звуковой информации

В основе кодирования звука в ПК лежит процесс преобразования колебаний воздуха в колебания электрического тока и последующая дискретизация

Аналоговый сигнал – это сигнал данных у которого каждый из параметров описывается функцией времени и непосредственно множеством возможных значений

Дискретизация – представление непрерывной функции в дискретной совокупности ее значений

Частота выборки (дискретизации) – обратная величина дескритизации

Цифровой звук – результат преобразования аналогового сигнала звукового диапазона в цифровой аудиоформат

Цифровая звукозапись – технология преобразования аналогового звука в цифровой

Теорема Котельникова:

Фундаментальное утверждение в области цифровой обработки сигналов связывающий непрерывные и дискретные сигналы и гласящее что любую функцию состоящую из частот от 0 до f_1 можно передавать с любой точностью при помощи чисел следующих друг за другом.

Виды звукозаписи:

- 1)Магнитная
- 2)Магнитострочная
- 3)Лазерная
- 4)Оптическая
- 5)На электронные носители

Виды сжатия:

- без потери данных
- с потерей данных

Аудиоредактор(волновой) – для редактирования звуковой информации в цифровом представлении.

Цифровая звуковая рабочая станция – система, предназначенная для записи хранения и редактирования воспроизведения аудио файлов.

Форматы без сжатия – wav,raw

Форматы с сжатием без потерь– flac,apple lossless

Форматы с потерями – mp2,mp3,wma

Формула для нахождения звуковой информации

$$N = 2^i$$

N – кол-во уровней сигнала

i – глубина звука

Объем звукового файла

$$V_{audio} = o * D * T * i$$

Видео – технология формирования записи, обработки и воспроизведения

Видеозапись – технология записи визуальной информации

Цифровое видео – совокупность технологий для взаимодействия с видео и т.д.

Стандарты передачи видео – NTSC, SECAM, PAL

VGA - 640 - 480

PAL - HD - 1366 - 768

FullHD - 1920 - 1080

...

Кадровая частота – кол-во сменяемых кадров за единицу времени

Разрешающая способность – способность устройства передавать мелкие детали изображения

Стандарт разложения – определяющая кол-во строк изображения, частоту смены кадров и режим развертки

Битрейт – кол-во битов в единицу времени

Качество видео – характеристика обработанного видео по сравнению с оригиналом

Цветовая субдискретизация – кодирование изображений со снижением цветового разрешения

Видео дорожка:

$$V_{image} = resolution * v * T * i$$

Размер видео файла:

$$V_{video} = V_{image} + V_{audio}$$

Лекция 5. Алгоритмы сжатия данных

Сжатие данных – алгоритмическое преобразование данных для уменьшения размера

Сжатия информации – процесс преобразования информации хранящейся в файле к виду при котором уменьшается избыточность в ее представлении и соответственно требуется меньший объем памяти для хранения

Степень сжатия:

$$K_C = \frac{V_C}{V_Q} 100\%$$

Метаданные – информация о другой информации

TNAME	FNAME	Name	CUnique
BOREHOLES	IDBH	Автономер скважины	
BOREHOLES	IDS	Пользовательский код скважины	✓
BOREHOLES	X	Координата X (восток)	
BOREHOLES	Y	Координата Y (север)	
GW_LEVELS	IDBH	Номер скважины	✓
GW_LEVELS	Date_M	Дата измерения уровня грунтовых вод	✓
GW_LEVELS	GW_Level	Глубина залегания уровня грунтовых вод	
GW_CHEM	IDBH	Номер скважины	✓
GW_CHEM	Date_S	Дата опробования	✓
GW_CHEM	Depth_S	Глубина отбора пробы	✓
GW_CHEM	Date_An	Дата анализа	
GW_CHEM	Pe	Нефтепродукты	

Классификация метаданных:

- 1) по содержанию
- 2) по отношению к ресурсу в целом
- 3) по возможности логического вывода

RLE алгоритм(кодирование повторово)

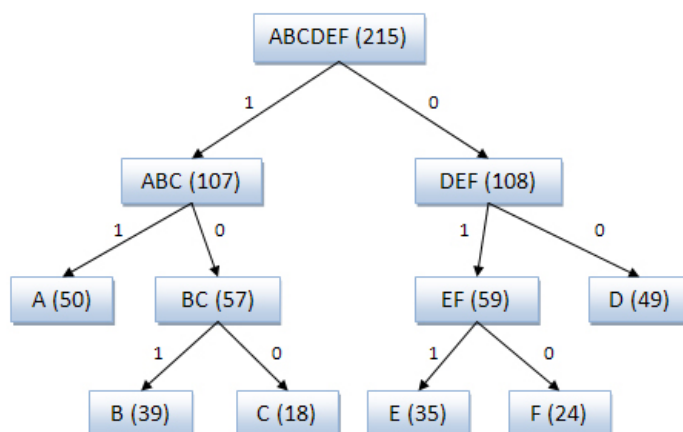
– алгоритм сжатия данных, заменяющий повторяющиеся символы (серии) на один символ и число его повторов. Серией называется последовательность, состоящая из нескольких одинаковых символов. При кодировании

(упаковке, сжатии) строка одинаковых символов, составляющих серию, заменяется строкой, содержащей сам повторяющийся символ и количество его повторов.

Алгоритм Шенона-Фано

- (
- 1) Символы первичного алфавита m_1 выписывают по убыванию вероятностей.
 - 2) Символы полученного алфавита делят на две части, суммарные вероятности символов которых максимально близки друг другу.
 - 3) В префиксном коде для первой части алфавита присваивается двоичная цифра «0», второй части — «1».
 - 4) Полученные части рекурсивно делятся и их частям назначаются соответствующие двоичные цифры в префиксном коде.
-)

Пример кодового дерева



Алгоритм Хаффмана

— жадный алгоритм оптимального префиксного кодирования алфавита с минимальной избыточностью. Был разработан в 1952 году аспирантом Массачусетского технологического института Дэвидом Хаффманом при написании им курсовой работы. В настоящее время используется во многих программах сжатия данных.

Этот метод кодирования состоит из двух основных этапов:
Построение оптимального кодового дерева.
Построение отображения код-символ на основе построенного дерева.

Классический алгоритм Хаффмана на входе получает таблицу частот встречаемости символов в сообщении. Далее на основании этой таблицы строится дерево кодирования Хаффмана (H-дерево).

1. Символы входного алфавита образуют список свободных узлов. Каждый лист имеет вес, который может быть равен либо вероятности, либо количеству вхождений символа в сжимаемое сообщение.
2. Выбираются два свободных узла дерева с наименьшими весами.
3. Создается их родитель с весом, равным их суммарному весу.
4. Родитель добавляется в список свободных узлов, а два его потомка удаляются из этого списка.
5. Одной дуге, выходящей из родителя, ставится в соответствие бит 1, другой — бит 0. Битовые значения ветвей, исходящих от корня, не зависят от весов потомков.
6. Шаги, начиная со второго, повторяются до тех пор, пока в списке свободных узлов не останется только один свободный узел. Он и будет считаться корнем дерева.

LZ77 и LZ78

— алгоритмы сжатия без потерь, опубликованные Абрахамом Лемпелем и Якобом Зивом в 1977 и 1978 годах соответственно. Эти алгоритмы стали основой других алгоритмов семьи LZ*: LZW, LZSS, LZMA и другие. Оба приведенных алгоритма используют словарный подход.

Алгоритм Лемпеля — Зива — Велча

— это универсальный алгоритм сжатия данных без потерь, созданный Авраамом Лемпелем, Яковом Зивом и Терри Велчем. Он был опубликован Велчем в 1984 году в качестве улучшенной реализации алгоритма LZ78, опубликованного Лемпелем и Зивом в 1978 году. Алгоритм разработан так, чтобы его было достаточно просто реализовать как программно, так и аппаратно

Более формально данный алгоритм можно описать следующей последовательностью шагов:

1. Инициализация словаря всеми возможными односимвольными фразами. Инициализация входной фразы W первым символом сообщения.
2. Если КОНЕЦ СООБЩЕНИЯ, то выдать код для W и завершить алгоритм.
3. Считать очередной символ K из кодируемого сообщения.
4. Если фраза WK уже есть в словаре, то присвоить входной фразе W значение WK и перейти к Шагу 2.
5. Иначе выдать код W , добавить WK в словарь, присвоить входной фразе W значение K и перейти к Шагу 2.