

## С о д е р ж а н и е   з а н я т и я

### Лекция 5    Информатика и ИКТ

*Подходы к измерению информации. Единицы измерения информации. Хранение информационных объектов различных видов на различных цифровых носителях. Определение объемов различных носителей информации. Архив информации.*

#### **Подходы к измерению информации**

Какое количество информации содержится, к примеру, в тексте романа "Война и мир", в картинах Рафаэля или в генетическом коде человека? Ответа на эти вопросы наука не даёт и, по всей вероятности, даст не скоро. **А возможно ли объективно измерить количество информации?** Важнейшим результатом *теории информации* является следующий вывод:

*В определенных, весьма широких условиях можно пренебречь качественными особенностями информации, выразить её количество числом, а также сравнить количество информации, содержащейся в различных группах данных.*

В настоящее время получили распространение два подхода к определению понятия "количество информации": **вероятностный** (чаще используется в теории информации) и **объемный** (чаще используется в вычислительно технике).

#### **Вероятностный подход**

**Вероятностный подход** предложил один из основоположников кибернетики американский математик Клод Шеннон. В качестве единицы информации он предложил принять один **бит** (англ. *bit* — *binary digit* — двоичная цифра).

**Бит** — количество информации, необходимое для различения двух равновероятных сообщений (типа "орел"—"решка", "чет"—"нечет" и т.п.).

Существует формула, связывающая *количество вариантов исхода  $N$*  и *количество информации  $I$* , которое несет сообщение:

$$N = 2^I$$

Следствие:  $I = \log_2 N$

**Пример 1:** *Сообщение об исходе опыта бросания монеты (2 равновозможных события — «орел» - «решка») содержит количество информации, равное одному биту.*

Действительно, возможных вариантов исхода  $N$  при бросании монеты — два "орел"—"решка". Значит,  $2 = 2^1$ . Следовательно  $I = 1$ .

**Пример 2:** *Какое количество информации содержит сообщение об исходе опыта бросания двух монет?*

**Решение:** Подсчитаем количество равновозможных исходов: орел-решка; решка-орел; орел-орел; решка-решка, т.е. четыре. Значит,  $N=4$ . Получим:  $4=2^2$

Отсюда,  $I=2$ .

**Ответ:** *Количество информации=2 бита.*

**Объемный подход**

В вычислительной технике **битом** называют наименьшую "порцию" памяти компьютера, необходимую для хранения одного из двух знаков: "0" и "1", используемых для внутримашинного представления данных и команд.

Бит — слишком мелкая единица измерения. На практике чаще применяется более крупная единица — **байт**, равная **восми битам**. Как мы уже отмечали (см. Лекцию 3) именно восемь битов требуется для того, чтобы закодировать любой из 256 символов алфавита клавиатуры компьютера ( $256=2^8$ ).

$$1 \text{ байт} = 8 \text{ бит}$$

**Единицы измерения информации**

Широко используются также ещё более крупные производные единицы информации:

$$1 \text{ Килобайт (Кб)} = 1024 \text{ байт} = 2^{10} \text{ байт},$$

$$1 \text{ Мегабайт (Мб)} = 1024 \text{ Кб} = 2^{20} \text{ байт},$$

$$1 \text{ Гигабайт (Гб)} = 1024 \text{ Мб} = 2^{30} \text{ байт}.$$

В последнее время в связи с увеличением объёмов обрабатываемой информации входят в употребление такие производные единицы, как:

$$1 \text{ Терабайт (Тб)} = 1024 \text{ Гб} = 2^{40} \text{ байт},$$

$$1 \text{ Петабайт (Пб)} = 1024 \text{ Тб} = 2^{50} \text{ байт}.$$

$$1 \text{ Эксабайт (Эб)} = 1024 \text{ Пб} = 2^{60} \text{ байт}.$$

Аналитики из Калифорнийского университета утверждают, что человечеству потребовалось 300 тысяч лет, чтобы создать первые 12 эксабайт информации, зато вторые 12 эксабайт были созданы всего за два года.

**Хранение информационных объектов различных видов на различных цифровых носителях**

Информация, закодированная с помощью естественных и формальных языков, а также информация в форме зрительных и звуковых образов хранится в памяти человека. Однако для долговременного хранения информации, ее накопления и передачи из поколения в поколение используются носители информации.

Материальная природа носителей информации может быть различной:

- молекулы ДНК, которые хранят генетическую информацию;
- бумага, на которой хранятся тексты и изображения;
- магнитная лента, на которой хранится звуковая информация;
- фото- и киноплёнки, на которых хранится графическая информация;
- микросхемы памяти, магнитные и лазерные диски, на которых хранятся программы и данные в компьютере, и так далее.

По оценкам специалистов, объем информации, фиксируемой на различных носителях, превышает один эксабайт в год. Примерно 80% всей этой информации хранится в цифровой форме на магнитных и оптических носителях и только 20% - на аналоговых носителях (бумага, магнитные ленты, фото- и киноплёнки).

Большое значение имеет надежность и долговременность хранения информации. Большую устойчивость к возможным повреждениям имеют молекулы ДНК, так как существует механизм обнаружения повреждений их структуры (мутаций) и самовосстановления.

Надежность (устойчивость к повреждениям) достаточно высока у аналоговых носителей, повреждение которых приводит к потере информации только на поврежденном участке. Поврежденная часть фотографии не лишает возможности видеть оставшуюся часть, повреждение участка магнитной ленты приводит лишь к временному пропаданию звука и так далее.

Цифровые носители гораздо более чувствительны к повреждениям, даже утеря одного бита данных на магнитном или оптическом диске может привести к невозможности считать файл, то есть к потере большого объема данных. Именно поэтому необходимо соблюдать правила эксплуатации и хранения цифровых носителей информации.

Наиболее долговременным носителем информации является молекула ДНК, которая в течение десятков тысяч лет (человек) и миллионов лет (некоторые живые организмы), сохраняет генетическую информацию данного вида.

Аналоговые носители способны сохранять информацию в течение тысяч лет (египетские папирусы и шумерские глиняные таблички), сотен лет (бумага) и десятков лет (магнитные ленты, фото- и киноплёнки).

Цифровые носители появились сравнительно недавно и поэтому об их долговременности можно судить только по оценкам специалистов. По экспертным оценкам, при правильном хранении оптические носители способны хранить информацию сотни лет, а магнитные - десятки лет.

### **Определение объемов различных носителей информации**

Носители информации характеризуются информационной емкостью, то есть количеством информации, которое они могут хранить. Наиболее информационно емкими являются молекулы ДНК, которые имеют очень малый размер и плотно упакованы. Это позволяет хранить огромное количество информации (до  $10^{21}$  битов в  $1 \text{ см}^3$ ), что дает возможность организму развиваться из одной-единственной клетки, содержащей всю необходимую генетическую информацию.

Современные микросхемы памяти позволяют хранить в  $1 \text{ см}^3$  до  $10^{10}$  битов информации, однако это в 100 миллиардов раз меньше, чем в ДНК. Можно сказать, что современные технологии пока существенно проигрывают биологической эволюции.

Однако если сравнивать информационную емкость традиционных носителей информации (книг) и современных компьютерных носителей, то прогресс очевиден:

- Лист формата А4 с текстом (набран на компьютере шрифтом 12-го кегля с одинарным интервалом) - около 3500 символов
- Страница учебника - 2000 символов
- Гибкий магнитный диск – 1,44 Мб
- Оптический диск CD-R(W) – 700 Мб
- Оптический диск DVD – 4,2 Гб
- Флэш-накопитель - несколько Гб
- Жесткий магнитный диск – сотни Гб

Таким образом, на дискете может храниться 2-3 книги, а на жестком магнитном диске или DVD - целая библиотека, включающая десятки тысяч книг.

## Архив информации

Созданную или полученную каким-либо образом информацию хранят в течение определённого времени, в течение которого её временно или долговременно содержат на различных носителях электронных данных. Если информация представляет интерес для её создателей или правообладателей, то им приходится создавать *электронные архивы*.

**Электронный архив** - это файл, содержащий один или несколько файлов в сжатой или несжатой форме и информацию, связанную с этими файлами (имя файла, дата и время последней редакции и т.п.).

Электронные архивы позволяют в любой момент времени извлекать из них необходимые данные для дальнейшего их использования в различных ситуациях (например, для обновления или восстановления утерянных данных). Такие архивы называют *страховочными копиями*. Их используют в случае утраты или порчи основной машиночитаемой информации, а также для длительного её хранения в месте, которое защищено от вредных воздействий и несанкционированного доступа. Как правило, компьютерными архивами информации являются электронные каталоги, базы и банки данных, а также коллекции любых видов электронной информации.

Для обеспечения надёжности хранения и защиты данных рекомендуют создавать по 2–3 архивные копии последних редакций файлов. В случае необходимости осуществляется разархивирование данных.

**Разархивирование** - это процесс точного восстановления электронной информации, ранее сжатой и хранящейся в файле-архиве.

Для создания архивных файлов и разархивирования используют специальные **программы-архиваторы**:

- WinRAR
- 7-Zip File Manager

**Основные возможности архиваторов:**

- просмотр содержания архива и файлов, содержащихся в архиве
- распаковка архива или отдельных файлов архива;
- создание простого архива файлов (файлов и папок) в виде файла с расширением, определяющим используемую программу-архиватор;
- создание самораспаковывающегося архива файлов (файлов и папок) в виде файла с пусковым расширением EXE;
- создание многотомного архива файлов (файлов и папок) в виде группы файлов-томов заданного размера (раньше - в размер дискеты).

**Контрольные вопросы:**

1. Какие существуют подходы к измерению информации?
2. Формула, связывающая *количество вариантов исхода  $N$*  и *количество информации  $I$* .
3. Дать определение *бита* при разных подходах к измерению информации.
4. Назвать основные единицы измерения информации (таблица).
5. Назвать цифровые носители информации, указать их достоинства и объёмы.
6. Назначение программ-архиваторов.