ЧАСТНОЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ «СЕВЕРО-КАВКАЗСКИЙ КОЛЛЕДЖ ИННОВАЦИОННЫХ ТЕХНОЛОГИЙ»

# ЛЕКЦИЯ 6. ТЕМА 2.3. ОСНОВНЫЕ ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И ИХ РЕАЛИЗАЦИЯ С ПОМОЩЬЮ КОМПЬЮТЕРОВ: ХРАНЕНИЕ, ПОИСК И ПЕРЕДАЧА ИНФОРМАЦИИ

## 1. Информационные процессы как предмет изучения информатики

Информационными процессами называют совокупность последовательных действий, выполняемых над информацией с целью её преобразования, накопления и использования. К фундаментальным информационным процессам относятся сбор, обработка, хранение, поиск и передача информации. Принципиальная особенность этих процессов состоит в том, что они могут быть реализованы как человеком, так и техническими устройствами, а информатика изучает их в первую очередь с позиций автоматизации средствами вычислительной техники.

Компьютер как универсальная машина для работы с информацией реализует все перечисленные процессы: обработка выполняется процессором, хранение — устройствами памяти и накопителями, передача — каналами связи, поиск — программными средствами доступа к данным. Три последних процесса объединяются общим свойством: они оперируют информацией как сохраняемым объектом, существующим во времени и пространстве, тогда как обработка имеет дело с её моментальным преобразованием. Рассмотрим хранение информации как базовый процесс, обеспечивающий существование всех остальных.

## 2. Хранение информации: сущность и структура данных

**Хранение информации** — процесс, в результате которого информация остаётся неизменной (содержательно тождественной) во времени на некотором носителе. Из этого определения следуют два аспекта проблемы хранения: физический (какая среда и каким способом удерживает записанные данные) и логический (как данные организованы, адресованы и защищены от искажения).

Любой информационный объект в компьютерной среде существует в виде последовательности двоичных разрядов, независимо от того, чем он является по содержанию — текстом, изображением, звукозаписью или исполняемой программой. Информационные объекты различаются типом, форматом представления и, как следствие, информационным объёмом. Типичные объёмные характеристики:

- текстовый документ в кодировке, отводящей один байт на символ, — от нескольких килобайт;
- фотография — единицы–десятки мегабайт в зависимости от разрешения и степени сжатия;
- минута звукозаписи несжатым потоком со стереофонической оцифровкой высокого качества (16 бит, 44,1 кГц, два канала) — около 10 МБ;
- час видеозаписи стандартной чёткости — от сотен мегабайт, видео сверхвысокой чёткости — десятки гигабайт;
- базы данных и программные комплексы — от единиц гигабайт до десятков и сотен гигабайт.

Отсюда видна масштабная лестница объёмов, в которой единицы измерения наращиваются степенями двойки: 1 Кбайт = $2^{10}$ байт, 1 Мбайт = $2^{20}$ байт, 1 Гбайт = $2^{30}$ байт, 1 Тбайт = $2^{40}$ байт, 1 Пбайт = $2^{50}$ байт. Терминологически следует различать биты и байты, а также учитывать различие двоичных приставок (Кбайт, МиБ) и десятичных (кБ, МБ), принятых в маркетинговой маркировке носителей: диск с маркировкой «1 ТБ» содержит $10^{12}$ байт, что соответствует примерно 931 Гбайт в двоичном исчислении.

## 3. Хранение информационных объектов на цифровых носителях

### 3.1. Классификация носителей

**Носитель информации** — материальная среда, используемая для записи и сохранения данных. По физическому принципу записи современные носители делятся на следующие классы.

**Магнитные носители.** Запись основана на перемагничивании ферромагнитного слоя. К классу относятся жёсткие магнитные диски (HDD), в которых данные сохраняются на вращающихся пластинах, и магнитные ленты. Достоинства магнитных носителей — высокая удельная ёмкость и низкая стоимость хранения одного байта; недостатки — механическая чувствительность и ограниченная скорость произвольного доступа. Современные HDD достигают ёмкости 36 ТБ (Seagate Exos M), а серийные диски на 40 ТБ — десять пластин по 4 ТБ — проходят тестирование с началом массового производства в 2026 году.

**Оптические носители.** Запись и считывание выполняются лазерным лучом, изменяющим отражательную способность рабочего слоя. Компакт-диски (CD) вмещают около 700 МБ, цифровые универсальные диски (DVD) — 4,7 ГБ в однослойном и 8,5 ГБ в двухслойном варианте, диски Blu-ray — 25 ГБ в однослойном исполнении и 50 ГБ в двухслойном. Формат BDXL доводит ёмкость до 100 ГБ (трёхслойная запись) и 128 ГБ (четырёхслойная запись). Существуют модификации дисков архивного класса (BD-R XL Archival Grade, M-DISC), позиционируемые как средства длительного хранения.

**Полупроводниковые (флеш-) носители.** Запись основана на сохранении заряда в ячейках энергонезависимой памяти (NAND). К классу относятся твердотельные накопители (SSD), флеш-карты и USB-накопители. SSD не содержат механических частей, обеспечивают на порядки более высокие скорости доступа, чем HDD, однако при прочих равных условиях дороже в расчёте на единицу ёмкости. Наибольшая ёмкость серийно выпускаемых SSD достигает 122,88 ТБ — это накопители корпоративного класса. Существенное свойство флеш-памяти — ограниченное число циклов перезаписи ячейки, что учитывается контроллерами изнашивания.

**Магнитные ленты и ленточные библиотеки.** Ленточные носители последовательного доступа остаются основным средством холодного хранения больших массивов данных благодаря рекордно низкой стоимости хранения и долговечности. Картриджи LTO-9 вмещают 18 ТБ (45 ТБ при сжатии), картриджи новейшего поколения LTO-10 — 30 ТБ (75 ТБ при сжатии); модификация LTO-10 с арамидной подложкой ленты доводит «сырую» ёмкость до 40 ТБ, или 100 ТБ при аппаратном сжатии, при скорости передачи несжатых данных до 440 МБ/с.

### 3.2. Сравнительные характеристики носителей

| Носитель | Тип доступа | Типичная ёмкость | Назначение |
|---|---|---|---|
| HDD 3,5" | произвольный | 8–36 ТБ | оперативные массивы данных, NAS |
| SSD | произвольный | 0,5–122,88 ТБ | системные и рабочие данные |
| USB-накопитель, флеш-карта | произвольный | десятки ГБ — 2 ТБ | перенос и резервирование |

| Носитель | Тип доступа | Типичная ёмкость | Назначение |
|---|---|---|---|
| Blu-ray (BD-R) | последовательный | 25 / 50 ГБ | дистрибуция, локальные копии |
| Blu-ray BDXL | последовательный | 100 / 128 ГБ | архивные копии |
| Магнитная лента LTO-10 | последовательный | 30–40 ТБ (до 75–100 ТБ при сжатии) nstor.ru +1 | холодное хранение, резервные копии |

**3.3. Логическая организация хранения**

Независимо от физического носителя данные организуются файловой системой — способом логического представления, поддерживающим именование, иерархическую структуру каталогов, адресацию, атрибуты доступа и контроль занятого пространства. Один и тот же информационный объект может быть сохранён на носителях различных классов, при этом меняется физическая форма представления, но не содержание. Именно эта инвариантность цифрового представления обеспечивает переносимость данных между носителями и является принципиальным отличием цифрового хранения от аналогового.

**4. Определение объёмов носителей информации**

Задача определения объёма носителя и требуемого места для хранения объектов решается в двух направлениях: от данных к носителю (вычисление необходимой ёмкости) и от носителя к данным (определение числа размещаемых объектов).

**Расчёт информационного объёма объекта.** Объём объекта вычисляется исходя из способа его кодирования. Для текста объём равен произведению числа символов на объём кода одного символа: $V=N\cdot i$, где $i$ — разрядность кода (при кодировке один байт на символ, например Windows-1251, $i=8$ бит; при UTF-16 — 16 бит и более). Для растрового изображения без сжатия $V=X\cdot Y\cdot i$, где $X\cdot Y$ — число пикселей, $i$ — глубина цвета. Для звукозаписи несжатого потока $V=f\cdot i\cdot k\cdot t$, где $f$ — частота дискретизации, $i$ — разрядность, $k$ — число каналов, $t$ — длительность.

**Определение фактической ёмкости носителя.** При переносе данных на носитель следует учитывать, что полезная ёмкость меньше номинальной. На маркетинговую маркировку распространяется десятичная система (1 ГБ = $10^9$ байт), тогда как операционная система отображает объём в двоичных единицах (1 ГиБ = $2^{30}$ байт ≈ 1,074 ГБ); кроме того, часть пространства занимают служебные структуры файловой системы. Так, диск Blu-ray объёмом 25 ГБ отображается в системе как около 23 ГБ полезного объёма.

**Пример.** Требуется определить, сколько текстовых документов объёмом 2 Кбайт размещается на карте памяти ёмкостью 8 ГБ. Номинальная ёмкость 8 ГБ = $8 \cdot 10^9$ байт ≈ 7,45 ГиБ; с учётом служебных затрат полезный объём близок к 7,3–7,4 ГиБ. Число документов определяется делением доступного объёма на объём одного документа: $7{,}4\cdot 230/2048 \approx 3{,}8\cdot 1067{,}4\cdot 230/2048 \approx 3{,}8\cdot 106$ объектов. Аналогичный расчёт для мультимедийных данных: на диск Blu-ray 50 ГБ помещается около двух часов видеозаписи высокой чёткости, тогда как на однослойный BD-R 25 ГБ — примерно два часа видео с разрешением 1080p при среднем сжатии.

**Прогнозирование потребности в ёмкости.** При проектировании хранилищ объёмы вычисляются суммированием приращений данных за период с учётом коэффициента избыточности резервирования. Практика централизованного хранения показывает, что

прирост объёмов данных опережает рост плотности записи носителей, что определяет непрерывную модернизацию хранилищ и спрос на носители предельной ёмкости: отраслевой прогноз предполагает появление HDD на 60 ТБ к 2028 году и преодоление порога 100 ТБ в 2030-х годах.

## 5. Архив информации

### 5.1. Понятие архива

**Архив информации** — организованная совокупность информационных объектов, предназначенная для долговременного сохранения с обеспечением возможности последующего поиска, извлечения и проверки целостности. В цифровом понимании архив представляет собой не пассивное хранилище файлов, а информационную систему, включающую сами объекты, метаданные (описательные, административные и структурные сведения о них), индексные структуры для поиска и механизмы контроля целостности.

Термин «архивирование» в информатике употребляется в двух значениях, которые следует различать. В широком смысле электронное архивирование (digital preservation) — совокупность стратегий и практик защиты электронных документов от устаревания и утраты информации, охватывающая весь жизненный цикл данных. В узком, технологическом смысле архивирование — это объединение одного или нескольких файлов в единый сжатый архивный файл средствами архиваторов (ZIP, RAR, 7z и др.), что решает задачи сокращения объёма, удобства переноса и объединения файловых групп.

### 5.2. Методы сокращения объёма

Сжатие данных, применяемое при архивировании, реализуется алгоритмами двух классов:

- **обратимое сжатие без потерь** (методы кодирования длин серий, словарные методы, статистическое кодирование Хаффмана) — используется для текстов, программ, таблиц, когда восстановление данных должно быть точным;
- **сжатие с частичной потерей информации** (алгоритмы JPEG, MPEG, аудиокодеки с потерями) — применяется к изображениям, видео и звуку, где психовизуальная избыточность допускает отбрасывание малозначимых данных.

Степень сжатия зависит от типа данных: текстовые и табличные данные сжимаются в несколько раз, а уже сжатые файлы и мультимедийные данные, закодированные алгоритмами с потерями, практически не уменьшаются повторно. В ленточных носителях аппаратное сжатие учитывается в маркировке: для LTO принят коэффициент 2,5:1, поэтому ёмкость указывается в двух значениях — нативном и сжатом.

### 5.3. Функции и организация электронного архива

Полноценный электронный архив обеспечивает выполнение функций, принципиально отличающих его от сетевой папки или облачного каталога: регистрация и классификация документов, ведение описей, индексация и поиск по атрибутам, разграничение прав доступа, контроль версий, фиксация истории изменений, резервное копирование и контроль целостности. В нормативной модели архивного дела РФ (Федеральный закон от 22.10.2004 № 125-ФЗ «Об архивном деле в Российской Федерации») к архивным документам предъявляются требования подлинности, неизменности, конфиденциальности и читаемости на протяжении установленных сроков хранения. elma365.com +1

### 5.4. Проблема долговременного хранения

Архивное хранение порождает специфическую проблему, не имеющую аналога в бумажном деле: **устаревание технологической среды**. Даже физически сохранный носитель со временем оказывается нечитаемым из-за снятия с производства аппаратуры и форматов. Этим определяются основные стратегии цифрового сохранения:

• **миграция** — периодический перенос данных с носителя на носитель с одновременной конвертацией форматов в актуальные стандарты; известна также как «рефрешинг» носителей;

• **репликация** — дублирование массива на нескольких носителях и, в надёжных системах, в географически распределённых площадках (правило «3-2-1»: три копии данных на двух типах носителей, одна из которых — за пределами основной площадки);

• **метаданные и эмуляция** — сохранение полной технической характеристики объектов, позволяющее в будущем воспроизвести их поведение в исходной программной среде.

Для долгосрочных архивов критическим параметром становится срок службы носителя. В этом отношении оптические носители архивного класса (M-DISC, BD-R Archival Grade) и магнитные ленты, выдерживающие десятилетия хранения при соблюдении температурно-влажностного режима, превосходят флеш-память, у которой хранение заряда в отключённой ячейке деградирует в течение нескольких лет.

Хранение информации образует основание всех информационных процессов: поиск и передача имеют смысл лишь постольку, поскольку данные сохранены и доступны. Современное хранение реализуется иерархией носителей — от высокоскоростных SSD до ленточных библиотек, — различающихся ёмкостью, стоимостью, скоростью доступа и долговечностью. Расчёт объёмов данных и носителей опирается на строгую систему единиц и методы кодирования, а учёт различий между номинальной и полезной ёмкостью носит прикладной характер. Архив информации при этом выступает не просто «большой папкой», а системой, соединяющей физическое хранение с логической организацией, поиском и защитой целостности — и именно эта система, а не отдельный носитель, гарантирует сохранность данных на десятилетия.