



Лекция 2

Числовые характеристики при
поиске информации и способы
хранения информации



Числовые характеристики при поиске ЭД в библиотеках и Интернете

- **Полнота поиска** – показатель, выражающий количественное отношение найденных релевантных сообщений к общему числу релевантных сообщений, содержащихся в том массиве, по которому велся поиск.

$$Q_{n.n.} = \frac{N_{н.р.д.}}{N_{р.д.}} 100\%$$



Числовые характеристики при поиске ЭД в библиотеке и Интернете

- **Точность поиска информационного сообщения** – показатель, выражающий отношение количества релевантных документов к общему числу документов, выданных информационной системой в ответ на запрос.

$$T_n = \frac{N_{p.д.}}{N_{н.д.}} 100\%.$$

Полнота - Точность

Полнота

В реальных системах невозможно достичь одновременно высокой полноты и точности. Поэтому при оценке результатов поиска используются комбинированные метрики.



Точность

Количественная мера информации

$$H(A) = - \sum_{i=1}^k p(a_i) \log_2 p(a_i)$$

$$H(A)_{\max} = \log_2 K$$

Сравнительный анализ различных носителей информации

Вид носителя	Объем хранимой информации	Надежность хранения информации
Лист бумаги формата А4	1-30 Кбайт	Невысокая. Срок жизни современной бумаги – 100 лет.
Дискета формата 3”	1,44 Мбайт	Невысокая. Целесообразно использовать при малом числе обращений. Кол-во циклов записи\чтения – 20
Жесткий диск	1 Тбайт 1956 – 10Мб, 1991- 1Гб, 2005 – 500 Гб, 2007 – 1 Тбайт	Высокая. Используется в рабочих станций при многократном обращении к информационным и программным ресурсам
Оптический диск	200-700 Мбайт	Высокая. Используется как в случае однократного, так и многократного процесса записи информации
DVD оптические диски	До 8 Гбайт	Высокая. Используется для записи видео и аудио информации. Срок хранения не более 30 лет
Магнито-оптические диски	До 640 Мбайт	Высокая. Используются для записи компьютерных программ различного назначения и для хранения личных цифровых данных
Zip накопители	До 250 Мбайт	Высокая. Используется для хранения больших объемов информации. Недостаток – высокая стоимость накопления и хранения информации.
Flash-память	64 Кбайт- 100 Гбайт	Высокая. Кол-во циклов записи\чтения – 10^5 . Используется в различных переносных системах и для перезаписи больших объемов информации.
Стриммеры	От сотен Гбайт до десятков Тбайт	Сравнительно высокая. Используются в электронных архивах.

Сравнительный анализ единиц накопления с объемным показателем ее носителя

Единица накопления	Объемный показатель
Килобайт- 10^3 байт	Одна машинопечатная страница – 2 Кбайта Фото невысокого качества – 100 Кбайт
Мегабайт – 10^6 байт	Фото высокого качества – 2 Мб Полное собрание Шекспира – 5 Мб Высококачественная музыкальная запись (10 мин) – 10 Мб
Гигабайт – 10^9 байт	Грузовик с книгами 1 Гб Хорошая коллекция из произведений Бетховена – 20 Гб Журнальная коллекция библиотеки – 100 Гб
Терабайт – 10^{12} байт	Университетская библиотека – 2 Тб Библиотека Конгресса США – 10 Тб БД Нац. центра исслед. климата – 400 Тб 240 тысяч песен, записанных в формате MP3
Петабайт – 10^{15} байт	Все университетские библиотеки США – 2Пб Емкость всех жестких дисков в 1995 г. – 20Пб

Способы сохранения информации





Мировое производство информации в 2003 г., сохраненное на различных носителях

Носитель информации	Нижние оценки, Тб	Верхние оценки, Тб	Относительная доля в общем объеме, %
Бумага	327	1 634	0,009 – 0,03
Пленка	7 700	420 000	0,24 – 7,5
Магнитные носители	3 400 000	5 200 000	92 – 99,7
Оптические носители	51	103	0,001 – 0,002
Всего:	3 408 000	5 600 000	100

Распределение печатной продукции США по видам изданий (2003 г.)

Вид печатной продукции	Верхняя граница, Тб	Относительная доля печатной продукции в общем объеме, %
Книги	5,5	0,08
Газеты	13,5	0,2
Научные журналы	1,6	0,02
Брошюры	0,3	0,005
Офисные документы	559	8,2
Почта	6 230	91,4



Распределение производства информации по различным видам оптических носителей (2003 г.)

Вид оптического носителя (СД)	Верхняя граница, Тб	Относительная доля вида носителя в общем объеме, %
Аудио CD	58	56
CD-ROM	1,1	1,0
DVD	43,8	43

Общие выводы по создаваемой и используемой информации:

1. Ежегодный мировой объем производимой информации увеличивается **экспоненциально**.
2. Доля информации на *цифровых* носителях с каждым годом увеличивается по сравнению с аналоговыми носителями.
3. Суммарное годовое потребление информации пользователями остается **постоянной величиной**.
4. При этом ежегодно увеличивается экспоненциальный разрыв между объемом производимой информации и ее потреблением пользователями.

Решения в среде Интернет:

1. Автоматизация средств поиска нужной конечному пользователю информации.
2. Формирование оптимальных алгоритмов поиска информации.
3. Разработка средств адресной доставки потенциально востребованной информации конечным пользователям.

Алгоритм выполнения задач с использованием компьютерных и телекоммуникационных технологий

