Some fundamental aspects of information storage and retrieval

Paul.Nieuwenhuysen@vub.ac.be
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium

Prepared for a tutorial at IIVO in Brussels, in September 2005, organised by IIVO
for delegates from Shangai, China, who are interested in the management of municipal archives

These slides should be available from the WWW site
http://www.vub.ac.be/BIBLIO/nieuwenhuysen/presentations/
note: BIBLIO and not biblio
- contents
- summary
- structure
- overview

of this tutorial

• Text information storage and retrieval systems
• Databases and computerized information retrieval
• Knowledge organisation: classifications and thesaurus systems
• Pictures on computers
• Disks for storage of information
• Compact Discs: CD and DVD for storage of information

- Interruptions
- Questions
- Remarks
- Discussions

are welcome
Text information
storage and retrieval systems

Database systems: definition

A database (management) system is a program or set of programs, providing a means by which a user can easily store and retrieve data in the form of “databases”.
Information retrieval software: related terms

• Software for information storage and retrieval (ISR software)
• Text(-oriented) database management systems (Text-DBMS)
• Text information management systems (TIMS)
• Document retrieval systems
• Document management systems

Information retrieval software: applications (Part 1)

• Archives
  Archived documents
• Documentation centres
  Documents
• Libraries
  Books / Documents
• Musea
  Objects / Books / ...
• Medical files
  Patient’s histories
• Marketing departments
  Clients / Potential clients
• Schools
  Courses / Teachers
• Bibliographic databases
  Publications / ...
Information retrieval software: applications (Part 2)

- Meeting calendars
- Product information
- Laboratories
- Personal documentation
- Patent office
- Co-operating information networks
- Meetings = conferences
- Product descriptions
- Recipes
- Documents
- Patents
- Documents / Persons / Institutes / Events / ...

Text-information management systems: characteristics and definition

The information in the database is text oriented. Therefore, several features are required:

» ability to store relatively long blocks of texts
» ability to retrieve items in which specific words or terms occur anywhere
Text-information management: from free-form to structure

Free form text information without structure

Text database with information structured in files, records, fields, sub-fields, with links/relations among records,...

Functions of database management software

• Input / edit using keyboard or batch input
• Indexing of the database(s)
• Browse / Search data from database
• Output
  (Sort / Display / Print to file / Print to paper)

+ • Export / Import
The various formats of records in a database

- Input format(s) = Edit format(s)
- Internal format for long term storage
- Format to facilitate retrieval = inverted file
- Display formats for output to display, printer, file
- Format for exchange/export purposes

Indexes in books and databases: a comparison

**Book**

- Index_term_1 page x1, y1, z1, ...
- Index_term_2 page x2, y2, z2, ...

**Database**

- Index_term_1 record nr. x1 / field type nr. x1 / field occurrence x1 / position x1
- record nr. y1 / field type nr. y1 / field occurrence y1 / position y1
- Index_term_2 record nr. x2 / field type nr. x2 / field occurrence x2 / position x2
- record nr. y2 / field type nr. y2 / field occurrence y2 / position y2

- Printed
- Invisible
Output from a database to various “devices”

- to video display
- to printer
- to computer file ("printing" to a file)

Hierarchy in the use of a database

Database structure
Input / Editing
Searching / Output
Security / privacy / protection of databases

- **Password for searching**
  specific database(s) and / or fields and / or record

- **Password for editing**
  specific database(s) and / or fields and / or records

- **Password for changing**
  » database structure
  » input and modification work sheets
  » sort and print formats of data in records
  » sort and print formats of records in a selection

Databases and computerized information retrieval
Information retrieval: via a database to the user

Information retrieval: the basic processes in search systems
Information retrieval systems: many components make up a system

- Any retrieval system is built up of many, more or less independent components.

- To increase the quality of the results, these components can be modified more or less independent of each other.

Information retrieval systems: important components

- the information content
- system to describe formal aspects of information items
- system to describe the subjects of information items
- concrete descriptions of information items = application of the used information description systems
- information storage and retrieval computer program(s)
- computer system used for retrieval
- type of medium or information carrier used for distribution
What determines the results of a search in a retrieval system?

1. the information retrieval system
   (= contents + computer system)

2. the user of the retrieval system
   and the search strategy applied to the system

A simple database architecture:
all records together form a database

The ‘sliced bread architecture’
» the bread is a “database”
» each slice of bread is a “database record”
» there are no relations between slices / records
» the retrieval system tries to offer the appropriate slices / records to the user
Layered structure of a database

- Database (File)
  - Records
    - Fields
    - Characters
  + in many systems: relations / links between records

Text retrieval and language: an overview

- Problems/difficulties related to language / terminology occur
- in the case of “multi-linguality”: “cross-language information retrieval”; that is when more than 1 language is used
  » in the contents of the searched database(s) and/or in the subject descriptors of the searched database(s) OR
  » in the search terms used in a query
- even when only 1 language is applied throughout the system
Text retrieval and language: a word is not a concept

Problem:
A word or phrase or term is *not* the same as a concept or subject or topic.

So, to ‘cover’ a concept in a search, to increase the recall of a search, the user of a retrieval system should consider an expansion of the query; that is: the user should also include other words in the query to ‘cover’ the concept.
Text retrieval and language: a word is not a concept

» synonyms!
(such as:
Latin names of species in biology besides the common names,
scientific names besides common names of substances in chemistry...)

Text retrieval and language: a word is not a concept

» narrower terms, more specific terms
(such as particular brand names);
including terms with prefixes
(for instance: viruses, retroviruses, rotaviruses...)
Text retrieval and language:
a word is not a concept: example

Example: Searching for the concept “sea” can or should involve for instance the following words in a
Boolean OR-combination:
baltic OR bay OR bays OR coast OR coastal OR coastline
OR coasts OR cove OR coves OR gulf OR mangrove OR
mangroves OR marine OR mediterranean OR noordzee OR
noordzeekust OR noordzeekusten OR ocean OR oceanic OR
oceans OR pacific OR reef OR reefs OR “saline-freshwater
interface” OR sea OR seas OR seashore OR seawater OR
seawaters OR shore OR shores

?? Question ??

Which problems in text retrieval are illustrated by the following sentences?
Time flies like an arrow. Fruit flies like a banana.

? 

Time flies like an arrow. Fruit flies like a banana.
Time flies like an arrow.
Fruit flies like a banana.

OK!

Text retrieval and language:
ambiguity of meaning

- Problem:
  A word or phrase can have more than 1 meaning.
  Ambiguity of the meaning of a word is a problem for retrieval.
  This decreases the precision of many searches.
  The meaning can depend on the context.
  The meaning may depend on the region where the term is used.
Text retrieval and language: ambiguity of meaning

• Example of a word:
  » Pascal the philosopher
  » Pascal the computer language

Text retrieval and language: ambiguity of meaning

• Example of sentences:
  » The banks of New Zealand flooded our mailboxes with free account proposals.
  » The banks of New Zealand flooded with heavy rains account for the economic loss.
Text retrieval and language: ambiguity of meaning

Problem:
Ambiguity of meaning may be the cause of low precision.

Text retrieval and language: ambiguity of meaning

• Method to solve the problem at the time of database production:
  » adding to each database record codes from a classification system or terms from a thesaurus system, and providing the user with knowledge about the system used;
  in some cases, this process is computerized (completely automatic or with intellectual intervention);
Text retrieval and language: ambiguity of meaning

- Method to solve the problem, provided by the computerized retrieval system:
  - offering to the user a partly computerized access to the subject description system and then linking to the database for searching

- searching normally (without added value), but adding value by categorizing the retrieved items in the presentation phase to assist in the ‘disambiguation’; this feature is offered for instance by
  - the public access module of the book catalogue of the library automation system VUBIS at VUB, Belgium, when a searching items that were assigned a particular keyword
Text retrieval and language: ambiguity of meaning

» Natural language processing of the queries:
  linguistic analysis to determine possible meanings of the query, which includes disambiguation of words in their context:
  “lexical” analysis = at the level of the word
  “semantic” analysis = at the level of the sentence
  However, most queries are short and therefore it is difficult to apply semantic analysis for disambiguation.

Text retrieval and language: ambiguity of meaning

» Natural language processing of the documents:
  linguistic analysis to determine possible meanings of a sentence, which includes disambiguation of words in their context:
  “lexical” analysis = at the level of the word
  “semantic” analysis = at the level of the sentence
  However, most retrieval systems do not apply this complicated method.
A word is not a concept
A concept is not a word

The most simple relation between words and concepts is NOT valid.

A concept cannot be “covered” by only 1 word or term; this may be the cause of low recall of a search.
The meaning of many words is ambiguous; this may be the cause of low precision of a search.
Text retrieval and language: relation with recall and precision

Recapitulating the two problems discussed, we can say that

• **Expansion** of the query allows to increase the *recall*.
• **Disambiguation** of the query allows to increase the *precision*.

A good text retrieval system solves some problems due to language

• accepts words / terms / phrases in the query of the user
• maps the words to corresponding concepts
• presents these concepts to the user who can then select the appropriate, relevant concept (“disambiguation”)
• searches for this concept, even in documents written in another language
• presents the resulting, retrieved documents in the language preferred by the user
Enhanced text retrieval using natural language processing

Text retrieval and language: conclusions

- The use of terms and language to retrieve information from databases/collections/corpora causes many problems.
- These problems are not recognized or underestimated by many users of search/retrieval systems = The power of retrieval systems is overestimated by many users.
- Much research and development is still needed to enhance text retrieval.
Hints on how to use information sources: Boolean combinations

Most text search systems understand the basic Boolean operators:

**OR**

= obtain records that contain one or both search terms

**AND**

= obtain records that contain both search terms

**NOT** or **ANDNOT** or **AND NOT**

= exclude records that contain a search term

In the case of computer-based information sources, use Boolean combinations of search terms when appropriate and when possible.
Hints on how to use information sources: Boolean queries

• Most text search systems understand the basic Boolean operators typed in capital characters:
  OR
  AND

• So this leads us to queries like for instance
  (word1 OR word2 OR word3 OR word4) AND (wordA OR wordB OR wordC)

Knowledge organisation:
classifications and thesaurus systems
Knowledge organisation: introduction

- To organise knowledge / documents / books / reports / information / data / records / things / items / materials for more efficient storage and retrieval, some related, similar tools / systems / methods / approaches are used.
- Often but not yet always, this process is assisted by a computer system.
- Good systems are expanded and updated when the need arises.
- The organization system applied should ideally be clearly and immediately visible or even searchable on computer, by the user of the materials.

Knowledge organisation: relations between tools

- Controlled vocabularies
- Thesauri
- Ontologies / Topic maps
Classification systems: introduction

Classification systems present the subjects in a logical order, usually going from the more general to the more specific.

Classifications systems: examples of universal systems

Universal means here: covering all subjects
Not just one but several competing systems exist.

Examples

» *Universal Decimal Classification = UDC* used mainly outside U.S.A.
» *Dewey Decimal Classification = DDC* used mainly in U.S.A.
» *Library of Congress Classification* used mainly in U.S.A.
» ...
**Thesaurus:**

description

- **Thesaurus (contents) =**
  - system to control a vocabulary
    (= words and phrases + their relations)
  - + the contents of this vocabulary

- **Thesaurus program =**
  - program to create, manage, modify and/or search a thesaurus using a computer

---

**Thesaurus relations**

Term(s) with broader meaning

\[ BT (= \text{Broader Term}) \]

\[ RT (= \text{Related Term}) \]

\[ UF (= \text{Use(d) For}) \]

\[ NT (= \text{Narrower Term}) \]

\[ Synonym(s) \]

Term

Other term(s)

Term(s) with narrower meaning
Thesaurus applications related to information searching (1)

• For producers of a database:
  To find/choose index terms to add these to items in a database, when terms are taken from a controlled vocabulary to increase precision and recall in the searches by users of the database.

Thesaurus applications related to information searching (2)

• For users of a database:
  When the database to be searched is produced with added descriptors (words and terms) that are taken from a controlled list of approved, selected words and terms, then the searcher can use some printed or computer-based system first, to find more and ‘correct’ suitable words and terms that belong to that controlled list of descriptors; then, the searcher can use these descriptors (and only these words or terms) in a database query.
Thesaurus applications related to information searching (3)

- For users of a database:
  When the database to be searched is NOT produced with added descriptors (words and terms) that are taken from a controlled list of words and terms, then the searcher can use one or several thesaurus systems first, to find more words and terms and more suitable words and terms; then the searcher can use these found words and terms to formulate a query for that database (to increase recall and precision).

Thesaurus systems that cover all subjects

- General systems
- Universal systems
- Covering all subjects
- Broad and shallow systems
- Horizontal systems
Thesaurus systems that cover all subjects: examples (1)

- Library of Congress Subject Headings (LCSH)
- thesaurus system built into word processing software
- thesaurus system that runs on a pc (independent of Internet)
  see for instance http://www.wordweb.co.uk/free/

Thesaurus systems that cover all subjects: examples (2)

- thesaurus systems that can be used through the WWW
  » http://thesaurus.plumbdesign.com/
General thesaurus system through the WWW: screenshot sea

General thesaurus system through the WWW: screenshot ocean
Thesaurus systems covering all subjects: comments

- An ideal, complete thesaurus that covers all subjects does not exist.

Thesaurus systems focused on a particular subject

- Focused on a particular subject domain = narrow and deep, vertical systems
Thesaurus systems focused on a particular subject: examples

- **ERIC**: education, information science, ...
- **Psychological Abstracts / PsycInfo**
- **Sociological Abstracts / SocioFile**
- **INSPEC**: physics, electronics, information technology
- **the Aquatic Sciences and Fisheries Information System**
- **Medline** (the Medical Subject Headings = MeSH)
- Various thesaurus systems for art and architecture can be found online: 
  [http://www.getty.edu/research/tools/vocabulary/](http://www.getty.edu/research/tools/vocabulary/)

A database of thesaurus systems is accessible online through [http://www.taxonomywarehouse.com/](http://www.taxonomywarehouse.com/)
Knowledge organization: classifications versus thesauri

- **Classification**
  - Good for placement of documents (because documents on many related subjects can be kept together)
  - Not well suited for computer searching (too complicated)

- **Thesaurus**
  - Not suited for placement of documents
    - (because documents with related subjects would NOT be kept together)
  - Well suited for computer searching
    - (relatively simple alphabetic listing of keywords)

Pictures on computers

Raster graphics versus vector graphics
Graphics formats: bitmaps and vector graphics

- Bitmap/raster graphics
  used in programs for “painting”

- Vector-based graphics
  used in programs for “drawing”

Graphics formats: bitmaps compared with vector graphics

<table>
<thead>
<tr>
<th>Property</th>
<th>Bitmap/Raster</th>
<th>Vector graphics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suited for “painting” and “retouching”</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Resolution adapted to output medium</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Size of files</td>
<td>large</td>
<td>small</td>
</tr>
</tbody>
</table>
### Graphics formats for bitmaps only

<table>
<thead>
<tr>
<th>File name extension</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>.BMP</td>
<td>Microsoft bitmap</td>
</tr>
<tr>
<td>.GIF</td>
<td>CompuServe / Unisys</td>
</tr>
<tr>
<td>.JPG</td>
<td>JPEG</td>
</tr>
<tr>
<td>.PCX, .PCC</td>
<td>Z-Soft Paintbrush</td>
</tr>
<tr>
<td>.PNG (PNG8, PNG24)</td>
<td>W3C</td>
</tr>
<tr>
<td>.RLE</td>
<td>Microsoft compressed bitmap</td>
</tr>
<tr>
<td>.TIF</td>
<td>Aldus PageMaker</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

### Graphics formats for vector graphics (+ bitmaps)

<table>
<thead>
<tr>
<th>File name extension</th>
<th>Name / Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR</td>
<td>Corel Draw!</td>
</tr>
<tr>
<td>CGM</td>
<td>Graphics Metafile</td>
</tr>
<tr>
<td>EPS</td>
<td>Encapsulated PostScript</td>
</tr>
<tr>
<td>IMG</td>
<td>GEM</td>
</tr>
<tr>
<td>PPT or PPS</td>
<td>Microsoft PowerPoint</td>
</tr>
<tr>
<td>SVG</td>
<td>Scalable Vector Graphics</td>
</tr>
<tr>
<td>WPG</td>
<td>WordPerfect Graphics</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Pictures on computers

Bitmap pictures

Graphics file formats for bitmap pictures: a comparison

<table>
<thead>
<tr>
<th></th>
<th>BMP</th>
<th>GIF</th>
<th>JPG</th>
<th>PNG24</th>
<th>TIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. no. of pixels per image</td>
<td>65535 x 65535</td>
<td>65535 x 65535</td>
<td>65535 x 65535</td>
<td>256<em>256</em>256 = 16 777 216</td>
<td>256<em>256</em>256 = 16 777 216</td>
</tr>
<tr>
<td>Multiple images in 1 file</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>more</td>
<td>more</td>
</tr>
<tr>
<td>Animation is possible</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Maximum number of colours in one image</td>
<td>256<em>256</em>256 = 16 777 216</td>
<td>256</td>
<td>256<em>256</em>256 = 16 777 216</td>
<td>more</td>
<td>256<em>256</em>256 = 16 777 216</td>
</tr>
<tr>
<td>Compress. of file size w/o loss of quality</td>
<td>LZW</td>
<td>JPEG</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>→ bits per pixel</td>
<td>RLE</td>
<td>RLE</td>
<td>RLE</td>
<td>RLE</td>
<td>RLE</td>
</tr>
<tr>
<td>+ (but colours can be lost)</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- BMP: RLE
- GIF: LZW
- JPG: JPEG
- PNG24: more
- TIF: more
Maximum number of colours displayed by computers: examples

- Examples of limitations / maximum number of colours on a computer display:
  » 8-bit = $2^{8} = 256$ typical for older systems
  » 16-bit = $2^{16} = 65536$ = “thousands of colors” = “High Color”
  » 24-bit = $2^{24} = (2^{8})(2^{8})(2^{8}) = 256*256*256$ = “True Color”
  » 32-bit = $2^{32}$

Disks
Disks: overview of various types

**Magnetic disks**
- Floppy disks
- Hard disks

**Optical disks**
- Laserdisc
- WORM
- CD-ROM
- DVD
- ...

Disks: comparison of types

<table>
<thead>
<tr>
<th></th>
<th>Floppy</th>
<th>Hard</th>
<th>Optical</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure rate</td>
<td>☺ High</td>
<td>☺ Low</td>
<td>☺ Low</td>
<td>☺ High</td>
</tr>
<tr>
<td>Disk capacity</td>
<td>☺ Low</td>
<td>☺ High</td>
<td>☺ Low</td>
<td>☺ Low</td>
</tr>
<tr>
<td>Storage cost per bit</td>
<td>☺ High</td>
<td>☺ Low</td>
<td>☺ Low</td>
<td>☺ Low</td>
</tr>
<tr>
<td>Speed of data access and transfer</td>
<td>☺ Lowest</td>
<td>☺ Low</td>
<td>☺ Low</td>
<td>☺ Low</td>
</tr>
<tr>
<td>Exchangeability</td>
<td>☺ '+'</td>
<td>☺ '+'</td>
<td>☺ '+'</td>
<td>☺ '+'</td>
</tr>
<tr>
<td>Transportability</td>
<td>☺ '+'</td>
<td>☺ '+'</td>
<td>☺ '+'</td>
<td>☺ '+'</td>
</tr>
<tr>
<td>Risk of disk crashes</td>
<td>☺ Low</td>
<td>☺ High</td>
<td>☺ Low</td>
<td>☺ Low</td>
</tr>
</tbody>
</table>

Disks: decreasing prices: 1995-…
Disks: formats

- Physical format
- Logical format
- Applications formats

Transfer rate

= the speed at which the computer reads data from a disk once the data is found (kB/s)

<table>
<thead>
<tr>
<th>Type</th>
<th>Transfer rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floppy disk</td>
<td>about 30 kB/s</td>
</tr>
<tr>
<td>CD-ROM</td>
<td>150 (digital audio, 1x) up to 8 000 kB/s (40x)</td>
</tr>
<tr>
<td>Hard disk</td>
<td>3 000 up to 30 000 kB/s</td>
</tr>
</tbody>
</table>
Evolution of disk capacities: advantages and disadvantages

- Floppy disks + Hard disks
  - 😊 Fast evolution: increasing capacity
  - 😞 Incompatibilities between floppy disk and disk hardware

- Compact disc
  - 😞 Stable from 1985: no increase of capacity
  - 😊 High compatibility with CD and DVD drives

Disk backup technologies: various systems

- Tape drives
- Removable drives
- Recordable CD (not rewritable)
- Recordable and rewritable CD
- Extra hard disk
- DVD-RAM
- Network drives
Disk backup technologies: tape drives

• +Pros:
  » inexpensive hardware
  » low media cost
  » large capacity

• -Cons:
  » slow
  » serial storage; no fast random access

Disk backup technologies: recordable CD (not rewritable)

• Named CD-R
• The files are not erasable; not rewritable
• +Pros:
  » inexpensive; low media cost
  » random access storage
  » disks can be read by most CD-ROM drives
• -Cons:
  » limited to 700 MB
  » Not erasable / rewritable / reusable
**Disk backup technologies: recordable and rewritable CD**

- **Named CD-RW**
- **The files are erasable; rewritable**
- **+Pros:**
  - inexpensive; low media cost
    (but more expensive than CD-R)
  - random access storage
- **-Cons:**
  - limited to 700 MB
  - disks can NOT be read directly by most CD-ROM drives

**Disk backup technologies: extra hard disk**

- **+Pros:**
  - fast
  - random access storage

- **-Cons:**
  - expensive but prices come down
  - not removable / cannot be kept off-site
Disk backup technologies:
DVD-RAM

- +Pros:
  - random access storage
  - large capacity

- -Cons:
  - drives are expensive
  - not many drives are available

Disk backup technologies:
DVD-RW and DVD+RW

- +Pros:
  - random access storage
  - large capacity
  - inexpensive

- -Cons:
  - fragile disks
  - low reliability
Compact Discs

Media based on optical discs

- WORM (not following the CD standards)
- CD-DA (digital audio)
- CD-ROM for PC, CD-ROM for Apple, ...
- [CD-i], [CDTV]
- Photo-CD
- CD-ROM - based WORM
  - = CD-R = CD-Recordable
  - or CD-RW = CD-Rewritable
- DVD…
Compact Disc = CD:
properties

- The dimensions are standardised:
  - the diameter is 120 mm,
  - and the hole in the centre is 15 mm
- Applied in CD-DA, CD-ROM, CD-XA, CD-i, Photo-CD,...
- Constant Linear Velocity (= CLV) while reading data
- Data stored on 1 side in a spiral
  - of valleys (named “pits”)
  - and plains (named “lands”)

Compact Discs:
reading the data

- Data are read in a drive with a laser beam which converts transitions pit-land to bits.
- The reader head does not touch the surface of the disc, so that there is no friction and thus no wear or risk of disc crashes.
- This medium offers random access to data like other disks for computers.
  - Thus access is fast in comparison with media that have to be read sequentially like tapes.
Compact Discs:
storage capacity (Part 1)

- 1 CD can store about 600 to 700 MB
  = 600 000 to 700 000 KB
- For comparison, we should realise that a common A4 sheet of paper can store an amount of information in the form of printed characters that would require about 2 kB of space on a computer.
- So one CD can store about the same amount of text information equivalent as 300 000 of such A4 sheets.

Compact Discs:
storage capacity (Part 2)

- 1000 paper sheets together make a pile of about 10 cm.

- So one CD corresponds in this view to a pile of about 30 m of paper sheets, which is a pile of paper as high as a 10-floor building.
CD-ROM: advantages in comparison with other information carriers (1)

- The formats are well standardised and the technology is stable; this ensures a high degree of compatibility.

- The information density is high.

- The cost of information storage per information unit is low.

- The disks are easy to store, to transport and to mail.

CD-ROM: advantages in comparison with other information carriers (2)

- Crashes with the reader head do not occur in normal use, and the disks resist well to wear.

- Random access to information is possible.

- CD-ROM systems are easy to use.
**DVD:**

**description**

- DVD is the official name; stood for Digital Video (or Versatile) Disc
- Standard for storing data on optical disc with a higher capacity than in the case of the older CD.
- First DVD discs and drives/players available in 1997.
- DVD drives/players read CD-ROM also.

---

**DVD-ROM:**

**comparison of design with CD-ROM**

<table>
<thead>
<tr>
<th>Feature</th>
<th>CD-ROM</th>
<th>DVD-ROM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disc diameter</td>
<td>120 mm</td>
<td>120 (or 80) mm</td>
</tr>
<tr>
<td>Data sides</td>
<td>1</td>
<td>1 or 2</td>
</tr>
<tr>
<td>Data layers</td>
<td>1</td>
<td>1 or 2</td>
</tr>
<tr>
<td>Tracks per inch</td>
<td>16 000</td>
<td>34 000</td>
</tr>
<tr>
<td>Laser wavelength</td>
<td>780 nm (infrared)</td>
<td>635 to 650 nm (red)</td>
</tr>
</tbody>
</table>
## DVD-ROM: comparison of performance with CD-ROM

<table>
<thead>
<tr>
<th>Feature</th>
<th>CD-ROM</th>
<th>DVD-ROM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>0.68 Gbyte</td>
<td>4.5 to 17 Gbyte</td>
</tr>
<tr>
<td>Data transfer rate</td>
<td>1.2 (up to 12) Mbit/s</td>
<td>&gt; 11 Mbit/s</td>
</tr>
</tbody>
</table>

---

Questions?  
Suggestions?  
Topics for discussion?