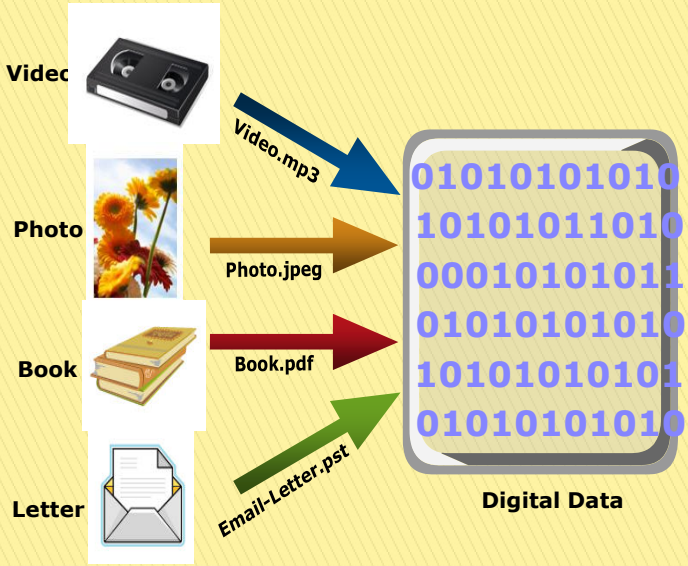
INFORMATION AND DATABASE MANAGEMENT SYSTEMS (IDBMS)

Instructor:
Dr. Sonam Nahar

Data

- Data is a representation of facts, concepts, or instructions in a formalized manner which should be suitable for communication, interpretation, or processing by humans or machines.
- Data is represented by characters such as alphabets, digits, or special characters.
- Most data is being converted into a digital format
  - Driven by user demand
  - Facilitated by
    - Increase in data processing capabilities
    - Lower cost and increased speed of storage
    - Affordable and faster Network



Who creates data?

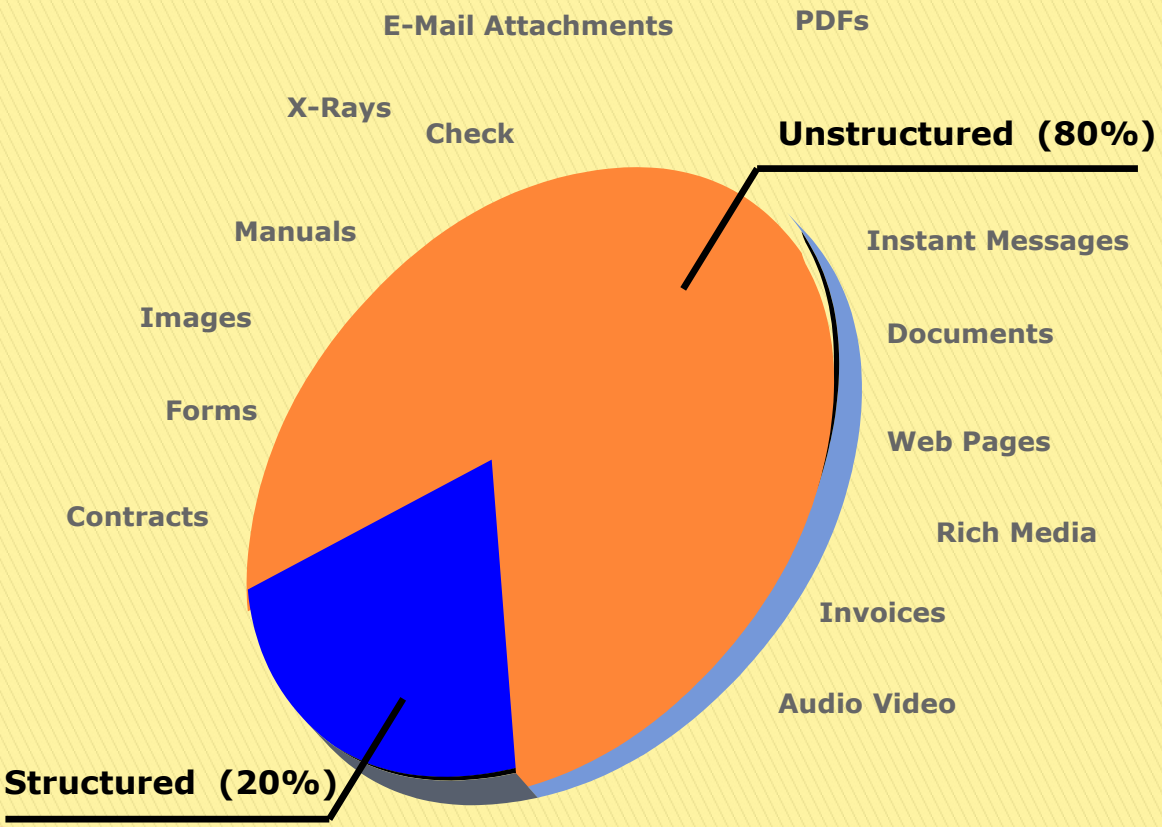
- Individuals
- Businesses

Categories of Data

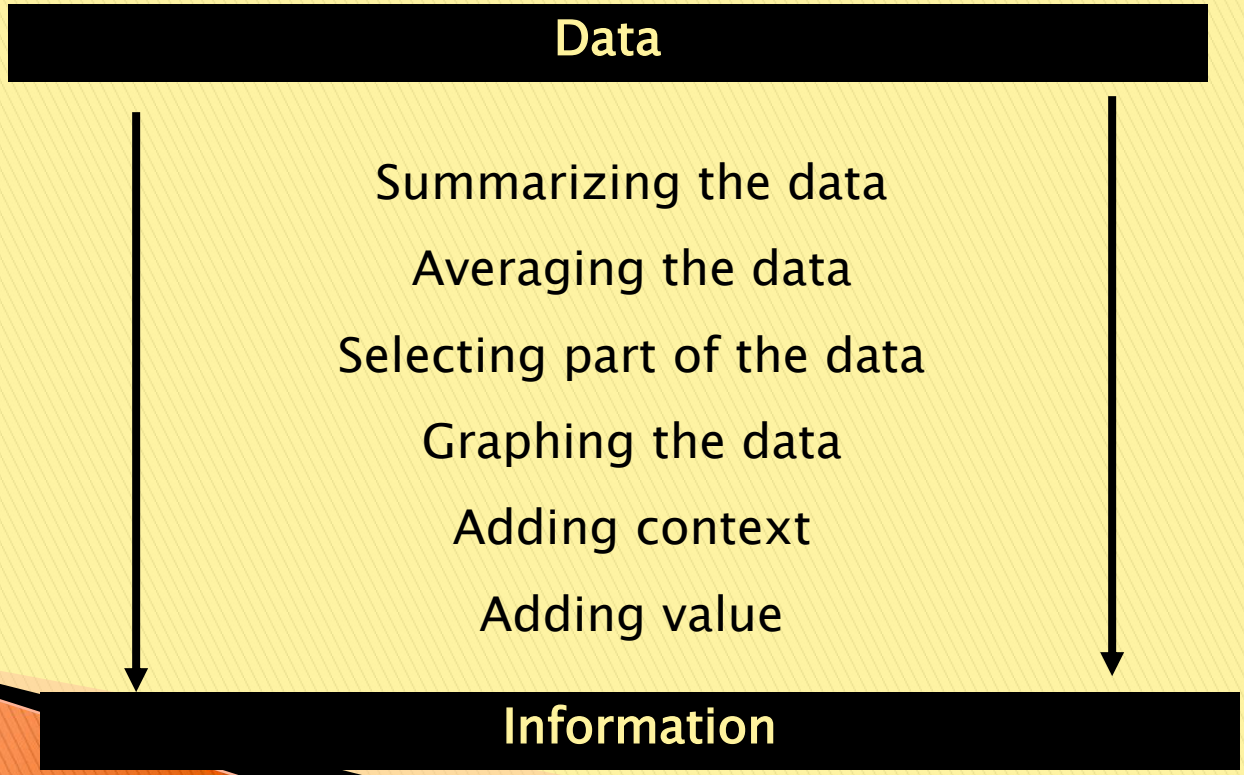
Data can be categorized as either structured or unstructured data

- **Structured**
  - Data Bases
  - Spread Sheets
- **Unstructured**
  - Forms
  - Images
  - Audio
  - Movies

Over 80% of Information is unstructured



Transformation of data into Information



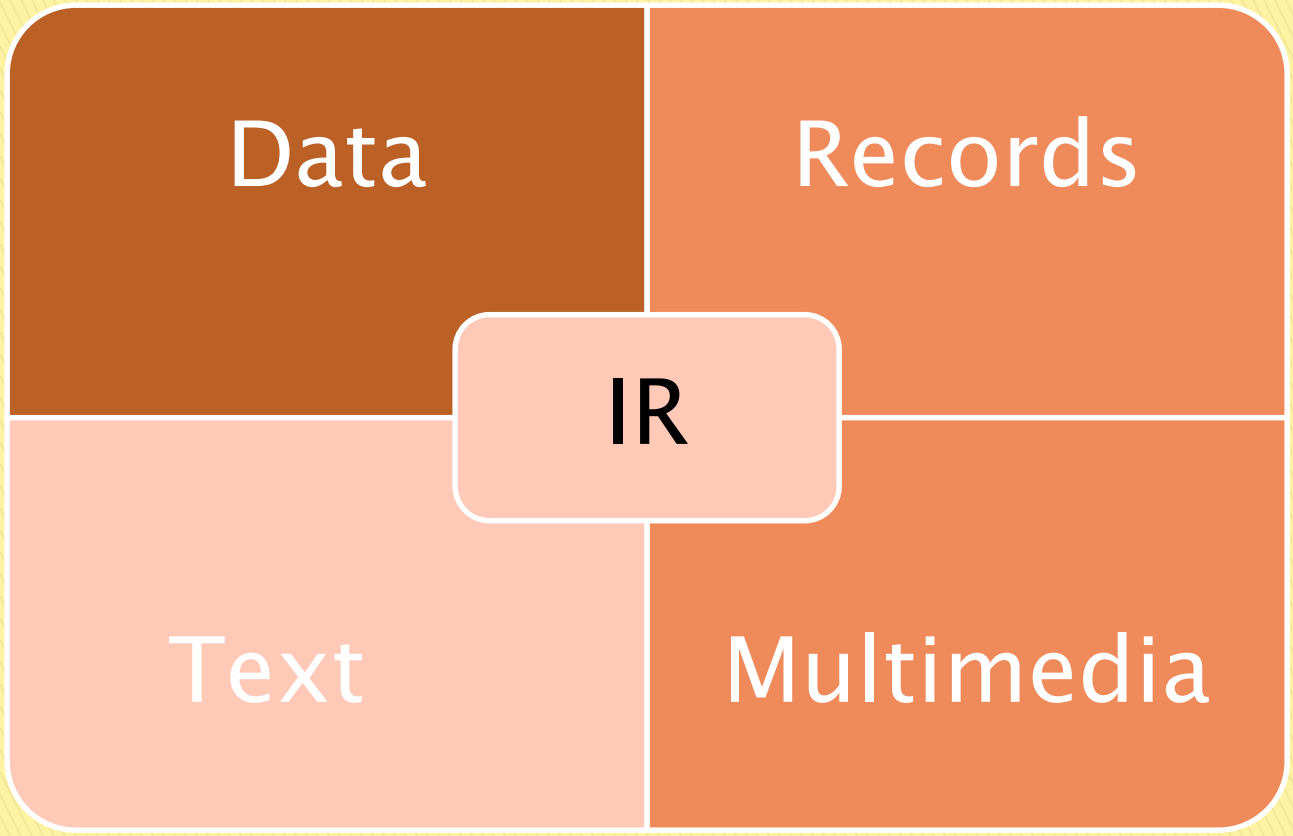
Information

Organized/Processed form of data is known as information

Definitions:

- data that have been processed so that they are meaningful;
- data that have been processed for a purpose;
- data that have been interpreted and understood by the recipient.
- data that have been processed on which decisions and actions are based.

Information Resources



Today Information Corpus

- **Files**

  Most sizeable companies have huge stores of electronic files scattered throughout the enterprise (a legacy of desktop networking). Letters, memos, reports, spreadsheets, database files, presentations, etc.

- **Databases**

  Companies usually maintain a number of databases on several different hardware and software platforms.

- **Email**

  Most employees communicate with email and much of an enterprise's internal and external business communication is done via email (and attachments).

- **Instant Messaging (IM)**

  This is becoming the way employees talk to one another in real-time.

- **Electronic Publishing**

  Most companies produce printed material such as catalogs, brochures, flyers, contact sheets, product specification sheets, newsletters, business reports, etc. Also, an increasing amount of information exists *only* in electronic format (e.g. Web pages, PDF documents, Intranets).

Why Information Storage?

"Digital universe – The Information Explosion"

- 21st Century is information era
- Information is being created at ever increasing rate
- Information has become critical for success

We live in an on-command, on-demand world

Example: Social networking sites, e-mails, video and photo sharing website, online shopping, search engines etc

Information management is a big challenge

Organization seek to *Store* *Protect* *Optimize*

Types of Information Representations

- Categories
- Equations
- Neural networks
- Natural language statements
- Logic statements
- Images

Information Capture

- Information capture is the process of collecting paper documents, forms and e-documents, transforming them into accurate, retrievable, digital information, and delivering the information into business applications and databases for immediate action
- Organizations and businesses need to determine the best way of carrying out data/information capture, as fits their purpose.

Information Capture

- Methods of information capture:
  - Manual method
  - Automated method
    - Optical character recognition (OCR)
    - Intelligent character recognition (ICR)
    - Optical mark reading (OMR)
    - Magnetic ink character recognition (MICR)
    - Smart cards
    - Web data capture
    - Voice recognition

Information Processing

There are many ways to apply the information stored in representations.

- Retrieval
  - Finding useful information
- Recognition
  - Identifying an instance
- Inference
  - Extend stored information to a new situation

Information Access

Humans gain access to information and data to support their needs by:

- Search
- Link
- Browse
- Navigate

Information Systems

- An organized system for the collection, organization, storage, and communication of information.
- An Information System is the system of persons, data records and activities that process the data and information in a given organization, including manual processes or automated processes.
- Information systems are created to capture, store, and support access of information representations.
- Information systems include the Web, databases, libraries, archives, and enterprise content management systems, etc.
- People use information systems so designers need to match content and the system interface to the user's needs.

Information Systems

- Computer based information system:

  A combination of

  Hardware

  Software

  Infrastructure and

  Trained personnel

  organized to facilitate

  Planning

  Control

  Coordination and

  Decision Making

  in an organization.

Components of Information Systems

Basically there are 5 components:

- Hardware

These are the devices like the monitor, CPU, and keyboard, all of which work together to accept, process, show data and information.

- Software

The term software refers to computer programs and the manuals (if any) that allow the hardware to process data

Components of Information Systems

- Data

Data are facts that are used by programs to produce useful information. Data is stored in files.

- Networks

Connecting system that allows diverse computers to distribute resources.

Components of Information Systems

- Procedures

Procedures are the policies that govern the operation of a computer system.

- People

Every system needs people if it is to be useful. Often the most overlooked element of the system are the people, probably the component that most influence the success or failure of information systems. This includes "not only the users, but those who operate and service the computers, those who maintain the data, and those who support the network of computers."

Types of Information Systems

- Executive Support System (ESS)

An Executive Support System ("ESS") is designed to help senior management make strategic decisions.

- Management Information System (MIS)

A management information system (“MIS”) is mainly concerned with internal sources of information and summarizes it into a series of management reports.

Types of Information Systems

- Decision Support System (DSS)

Decision-support systems ("DSS") are specifically designed to help management make decisions in situations where there is uncertainty about the possible outcomes of those decisions.

- Knowledge Management System (KMS)

  Knowledge Management Systems ("KMS") exist to help businesses create and share information.

Types of Information Systems

- Transaction Processing System (TPS)

  Transaction Processing Systems ("TPS") are designed to process routine transactions efficiently and accurately.

- Office Automation System (OAS)

  Office Automation Systems are systems that try to improve the productivity of employee who need to process data and information.

Information Management

- **Information management** is the management of organizational processes and systems that acquire, create, organize, distribute, and use information.
- IM is a continuous cycle of six closely related activities:
  1. identification of information needs;
  2. acquisition and creation of information;
  3. analysis and interpretation of information;
  4. organization and storage of information;
  5. information access and dissemination;
  6. information use

Information Management Systems

- Information Management:
  - Management of information resources.
  - Design of information technology components.
  - Analysis of information processing procedures.
  - Deriving knowledge from the *information corpus*.
- Information Management System:

  A general term for software designed to facilitate/manage the storage, organization, and retrieval of information.

Roles Associated with Information Management

- Database administrator
- Data modeler
- Data Analyst
- Application developer
- End user

Goals of IM

1) Supply work, business and consumption processes with information — This is the basic goal: work cannot be done without required information.

2) Improve and speed up business, work and consumption processes through information use and efficient information processing — Information is not only one of the inputs to the work process. By improving information supply and its processing, the whole process usually can be made more efficient.

3) Create and maintain competitive advantage through new, IT-based work and business processes — Often, information technologies allow reorganization of work in completely new ways, and creation of totally new businesses.

Goals of IM

4) Efficient use of organization's information assets — While previous goals come from activity (process), this goal statement invites to think about organization's information not as some side-product of activity, but as the central resource. Information, not activity may be the „real thing".

5) Reduce unnecessary complexity of information processing systems; protect against information overload.

Information Storage & Retrieval

Information Storage and Retrieval

Systematic process of collecting and cataloging data so that they can be located and displayed on request. Computers and data processing techniques have made possible to access the high-speed and large amounts of information for government, commercial, and academic purposes.

Information Storage

Storage keeps data and information for use in the future. Common storage mediums are:

1. Hard Drive
2. Floppy Disk
3. CD&DVD
4. USB Flash Drive
5. Cloud

Information Retrieval (IR)

- An information retrieval system is an information system, that used to store items of information that need to be processed, searched, retrieved, and disseminated to various user populations.
- Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for metadata that describe data, and for databases of texts, images or sounds.

Information Storage and Retrieval Systems

1. Document retrieval systems
2. Database systems
3. Reference retrieval systems
4. Multimedia retrieval systems

Information Representation

- Information needs to be represented before it can be retrieved.
- Information representation includes the extraction of some elements (e.g., keywords or phrases) from a document or the assignment of terms (e.g., descriptors or subject headings) to a document so that its essence can be characterized and presented.
- Typically, information representation can be done via any combination of the following means: abstracting, indexing, categorization, summarization, and extraction.

Major Components of IR

Information retrieval can be divided into several major constitutes which include:

1. Database
2. Search mechanism
3. Language
4. Interface

Major Components of IR

Database:

- A system whose base, whose key concepts, is simply a particular way of handling data & its objective is to record and maintain information.
- Databases in IR comprise information represented and organized in a certain manner. In the traditional sense, a database (e.g., an online database) is typically made of records that can be further decomposed into fields, the smallest and most natural units for sorting, searching, and retrieving information.

Major Components of IR

Search Mechanism

- Information organized systematically that can be searched and retrieved when a corresponding search mechanism is provided.
- Search procedures can be categorized as basic or advance search procedure.
- Basic search procedures are commonly found in the majority of operational IR systems, while advanced search procedures have been tested and experimented with mainly in laboratories. However, in recent years, advanced search algorithms have increasingly been integrated into internet retrieval systems.
- Capacity of search mechanism determines what retrieval techniques will be available to users and how information stored in databases can be retrieved.

Major Components of IR

Language

- Information relies on language when being processed, transferred or communicated.
- Language can be identified as natural language and controlled vocabulary.

Interface

Interface regularly considered whether or not an information retrieval system is user friendly.

- Quality of interface checked by interaction mode (e.g., menu selection), display features (e.g., screen layout and font type), and other related factors.
- Determines the ultimate success of a system for information retrieval

Evaluation of IR Systems

We can evaluate an information retrieval system by considering the following three issues.

- How well the system is satisfying its objectives, how well it is satisfying the demands placed upon it
- How efficiently it is satisfying its objectives and finally
- Whether the system justifies its existence

Evaluation Measures for Information Retrieval

Recall and Precision

Measure of whether or not a particular item is retrieved or the extent to which the retrieval of wanted items occurs

- The performance of a system is often measured by **recall** ratio, which denotes the percentage of relevant items retrieved in a given situation.

$$\text{Recall} = \frac{\text{Number of relevant items retrieve}}{\text{Total number of relevant items in the collection}} \times 100$$

Evaluation Measures for Information Retrieval

- By **precision** we mean how precisely a particular system functions. It is quite obvious that when the system retrieves items that are relevant to a given query it also retrieves some documents that are not relevant

$$\text{Precision} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of items retrieved}} \times 100$$

Course Structure

Credits: 4

40 lectures and 10 labs

Course Outcomes:

- Understand the different issues involved in the design and implementation of a database system.
- Apply the modeling concepts and notation of the relational data model.
- Determine database storage structures and access techniques for a given problem.
- Understand the basic working of database management aspects in terms of transaction processing, concurrency control, and recovery.

Syllabus:

1. Information Management Concepts
2. Introduction to DBMS
3. Data Modeling
4. Relational Databases
5. Query Languages
6. File structures, Indexing, and Hashing
7. Transaction Processing, Concurrency Control, and Recovery
8. Advanced Topics

Text Book References:

- Silberschatz, H. Korth, and S. Sudarshan, *Database System Concepts*, McGraw-Hill.
- R. Elmasri and S. Navathe, *Fundamentals of Database Systems*, Addison-Wesley, 6th ed., 2011
- R. Ramakrishnan, *Database Management Systems*, WCB/McGraw-Hill.
- C.J. Date, An Introduction to *Database Systems*, Pearson, 8th ed.

Text Book References:

- R. Elmasri and S. Navathe, *Fundamentals of Database Systems*, Addison-Wesley, 6th ed., 2011
- Silberschatz, H. Korth, and S. Sudarshan, *Database System Concepts*, McGraw-Hill.
- R. Ramakrishnan, *Database Management Systems*, WCB/McGraw-Hill.
- C.J. Date, An Introduction to *Database Systems*, Pearson, 8th ed.

Grading Policy:

Component	Weightage
Lecture	
Continuous evaluation (Quiz, Assignment)	30%
Midterm	30%
End Term Examination	40%
Lab	
Continuous Evaluation/Assignments/Attendance	30%
Midterm	35%
End-term	35%