Data Processing

by

Assoc. Prof. Pannipa Phaiboonnimit

Adapted for English Section by Kittipitch Kuptavanich

Data Processing

- **A process which summarizes, analyzes or otherwise converts data into usable information.**

| Raw data | → | Process | → | Information |

Raw Data

- **Data collected from a source and are the data input to processing**

- **Example:**
    - **Numbers, texts, pictures, sound, temperature, velocity**
    - **A group of students' exam scores**

Information

- **Information is the end product of data processing**

- **Example:**

  - **The average of a group of students' exam scores**

  - **Weather forecast from weather data**

Type of Data Processing

- **Manual Data Processing**

  - **Performed by people**

- **Automatic Data Processing**

  - **Online Processing**

  - **Batch Processing**

Online Processing

- **Real-time Processing**

  - **Airline Reservation System**

- **Direct access media is used**

- **Immediate output**

- **Up-to-date Information**

- **I/O device connected with CPU**

- **Example: Airline Ticketing**

Batch Processing (1)

- **The data is collected over a period of time**

- **Processing will done with a collection (a batch) of data**

- **I/O device not connected with CPU**

- **Data is not up-to-date**

- **Cheaper to process**

- **Example: Grading by computer, Billing**

Batch Processing (2)

- Master Files
  - The actual updating of master files happen periodically
  - Inventories of goods

- Transaction Files
  - Record changes to be updated to the master files
  - List of items bought/sold during the day

Electronic Data Processing

Data Collection → Data Processing → Information

Data Collection

- Collection Method

- Characteristics of Good Data

- Data Encoding

- Verification and Validation

Data Collection Method

- **Observation**
  - **Example: counting number of cars passing through an intersection.**
- **Measurement**
  - **Example: set up a pollution sensor and read data from it.**
- **Document**
  - **Example: Go through existing records in the library.**
- **Interview**

What is Good Data?

- **Accurate**

- **Recent**

  - **Is the data up-to-date**

- **Complete**

  - **Does the data cover everything we intend to collect?**

Data Encoding

- **Save data storage space**

- **Faster data access**

- **Use number and code instead of text**

- **Using different unit for number**

- **Good for fields that repeat often**

  - **Department, province, country**

Example: Data Collected From Questionnaire

Database Coding Scheme

ID		Sex		
		Male(M)	Female(F)	
Name		Lastname		
Major				
1. English	2. Spanish	3. German	4. Chinese	5. etc
Income				
1. < 2500	2. 2501-5000	3. 5001 – 7500	4. 7501 – 10K	5. > 10K

Data Encoding

ID	Sex	Name	Last Name	Major	Income
33020	Male	John	Smith	Spanish	11,000

ID	Sex	Name	Last Name	Major	Income
33020	M	John	Smith	2	5

Saving Space with Encoding

- For example, province:

C	h	i	a	n	g	-	M	a	i	-	-	-	-	-	-	-	-	-	-

<div align="right">20</div>

- Now, with encoded by 2-digit number:

0	4

- Also, harder to make/easier to detect mistakes

Saving Space with Encoding (cont.)

- Although, you will also need an encoding table to translate which shorten code mean which full data. You will still save space for large amount of entries.

| Code	Province
01	Amnat Charoen
02	Bangkok
...	
76	Yasothon

Verification and Validation

- **Make sure that your data is correct**

- **Verify ─ Is the data entered the same as data collected?**

- **Validate ─ Does the data make sense?**

  - **Range Check: Is the field value in reasonable range?**

    - **Age of 300? -5?**

  - **Relation Check: Does the value make sense, considering the other related field values?**

    - **A women with *Mr.* as title?**

- **Edit if there are errors**

Database Diagram

Data Organization

BIT
CHARACTER
FIELD
RECORD
FILE
DATABASE
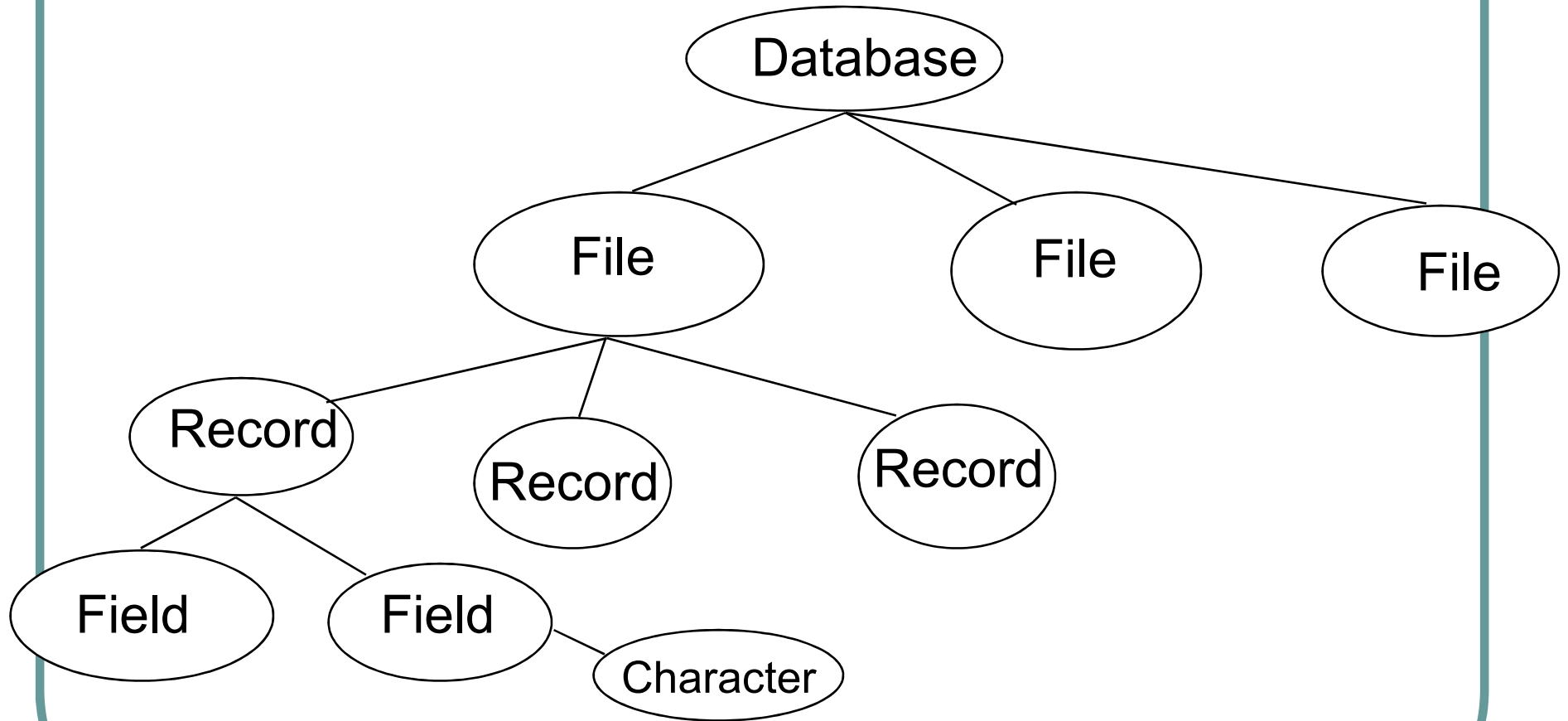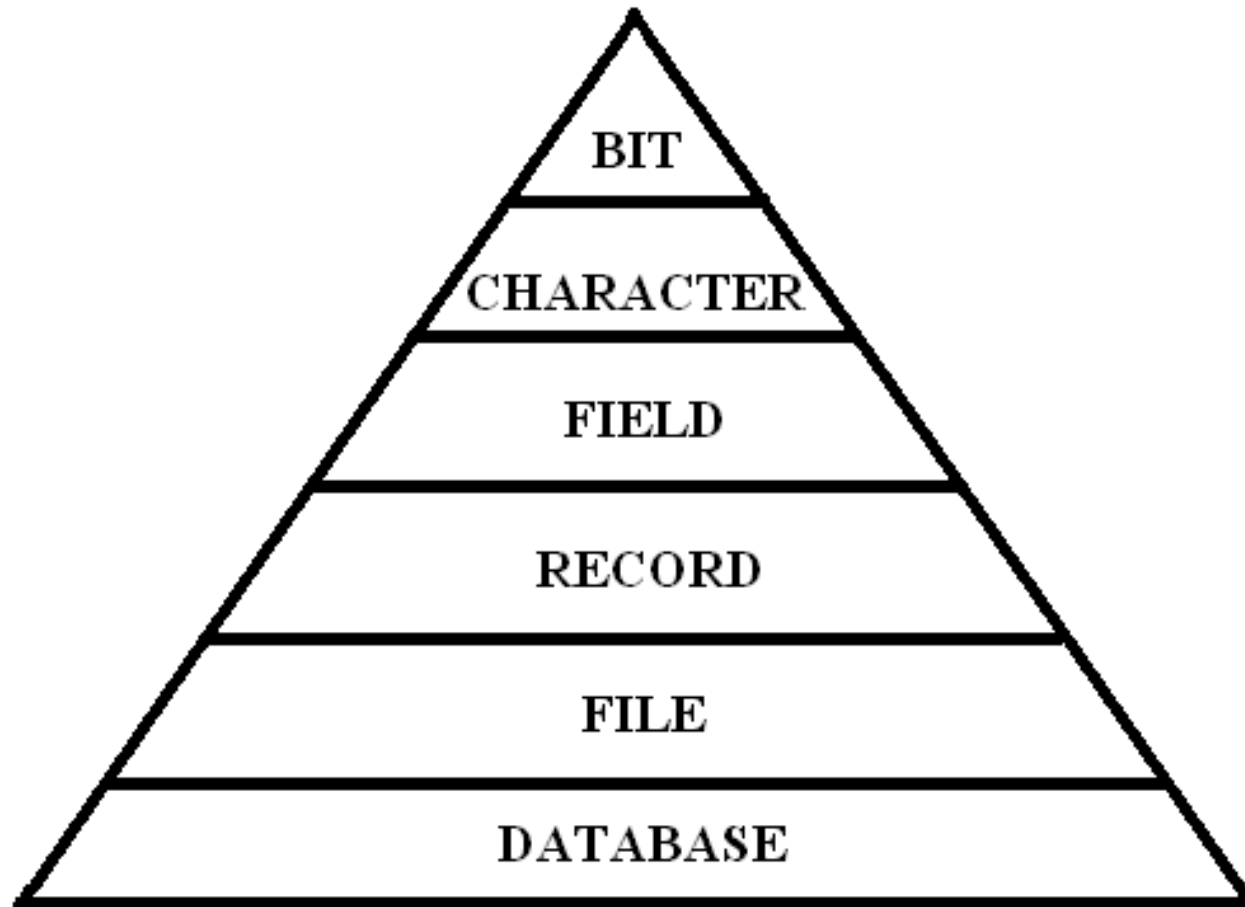
Database

- A data collection in a structured and standard form

- One set of DB for the whole organization

- Data Independent

- Can be accessed using Query Language

File

- A collection of one entity

- 3th GL of data collection

- File is data dependent (program must know the structure of a file)

- Tape, disk, CD ROM are data media

File, Record, Field

Dept	EmployeeID	Name	Hour	Rate
Finance	55028	John Smith	35	8.75
Finance	54019	Jane Smith	40	11.25
Research	56023	Rock Lee	37	9.5
Research	50017	Jackie Chan	40	8.75

Fields

File

Record

Database Example

Dept	EmployeeID	Name	Hour	Rate
Finance	55028	John Smith	35	8.75
Finance	54019	Jane Smith	40	11.25
Research	56023	Rock Lee	37	9.5
Research	50017	Jackie Chan	40	8.75

Department	Employee#	Starting Time
Finance	15	8:00
Research	22	12:00

Data Processing

- Calculating

- Updating

- Reporting

- Summarizing

- Sorting

- Searching

- Classifying

Calculation

- From existing attributes (fields), find:

  - Mean

  - Median

  - Mode

Updating

- Calculate and replace the values of existing attributes

    - Raising employees' salary

    - Adding interest to bank account balance

Reporting

Dept	EmployeeID	Name	Hour	Rate	Payment
Finance	55028	John Smith	35	8.75	306.25
Finance	54019	Jane Smith	40	11.25	450
Research	56023	Rock Lee	37	9.5	351.5
Research	50017	Jackie Chan	40	8.75	350

Summarizing

Dept	EmployeeID	Name	Hour	Rate	Payment
Finance	55028	John Smith	35	8.75	306.25
Finance	54019	Jane Smith	40	11.25	450
Research	56023	Rock Lee	37	9.5	351.5
Research	50017	Jackie Chan	40	8.75	350

Average= 364.44

Sorting

- Put records in a particular order

  - By number (grade, salary, etc.)

  - By text (name)

  - By ID

Searching

➢ Find records that match a particular criteria

● "Student with GPA lower than 1.75"

● "Employee in Finance department OR with salary higher than 20,000"

● "Student with first name of John AND with GPA higher than 3.00"

Classifying

➢ Split records into related groups

● Divide students by majors

● Divide employees by departments

Computer Data Processing w/ Feedback

● Processing
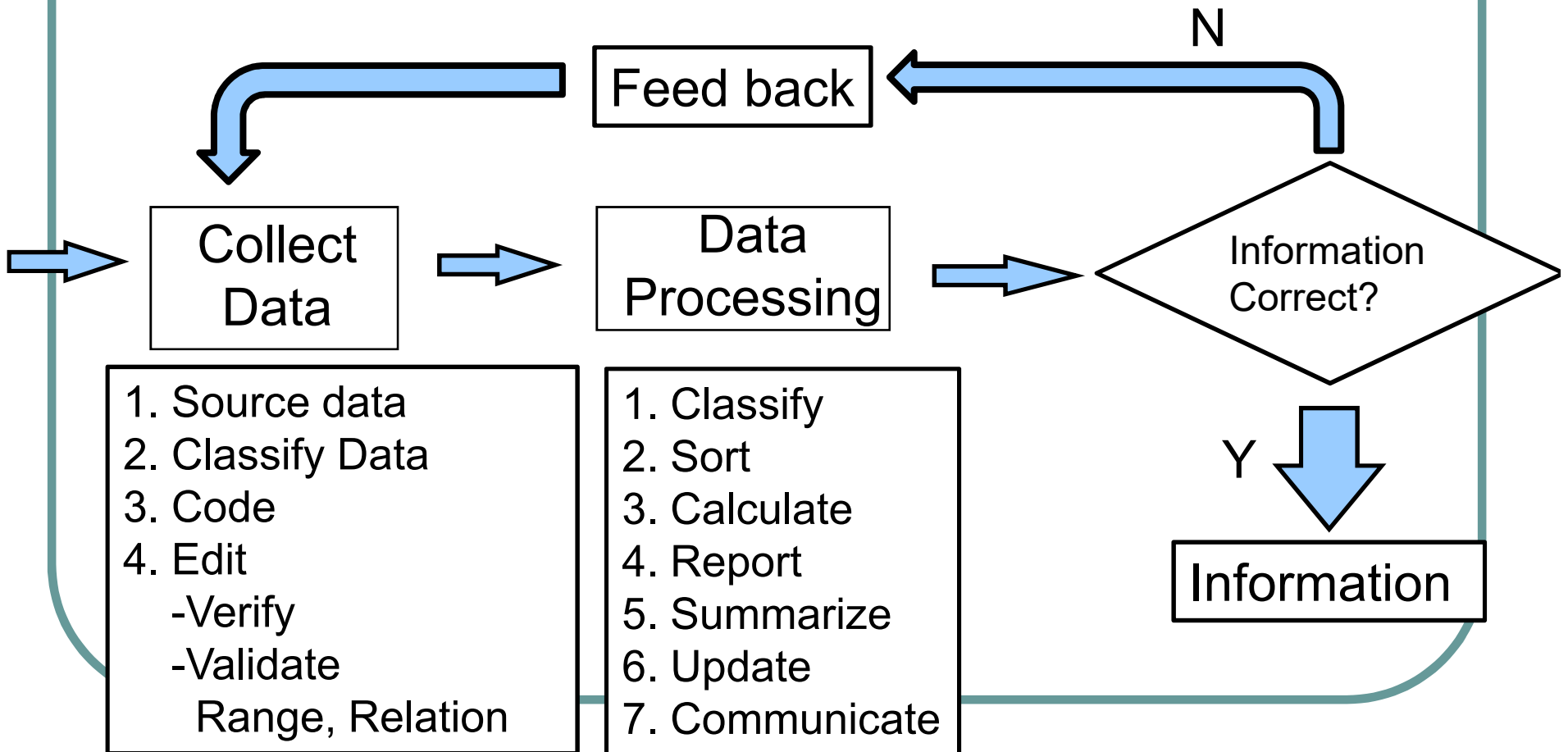
Example: Expense Tracking

Data Collection	→	Process	→	Information

Expenses

- "Lunch ฿35"

- **Classifying**: divide expense by dates/categories
- **Calculating**:
  - Calculating daily expense

Expenses by date/category

- "I've spent ฿2,500 on food this month."