RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-04

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- KNOWLEDGE DISCOVERY IN DATABASES (KDD)
- ARCHITECTURE OF DATA MINING SYSTEM
- DATA MINING SYSTEM CLASSIFICATION
- MCQ
- REFERENCES

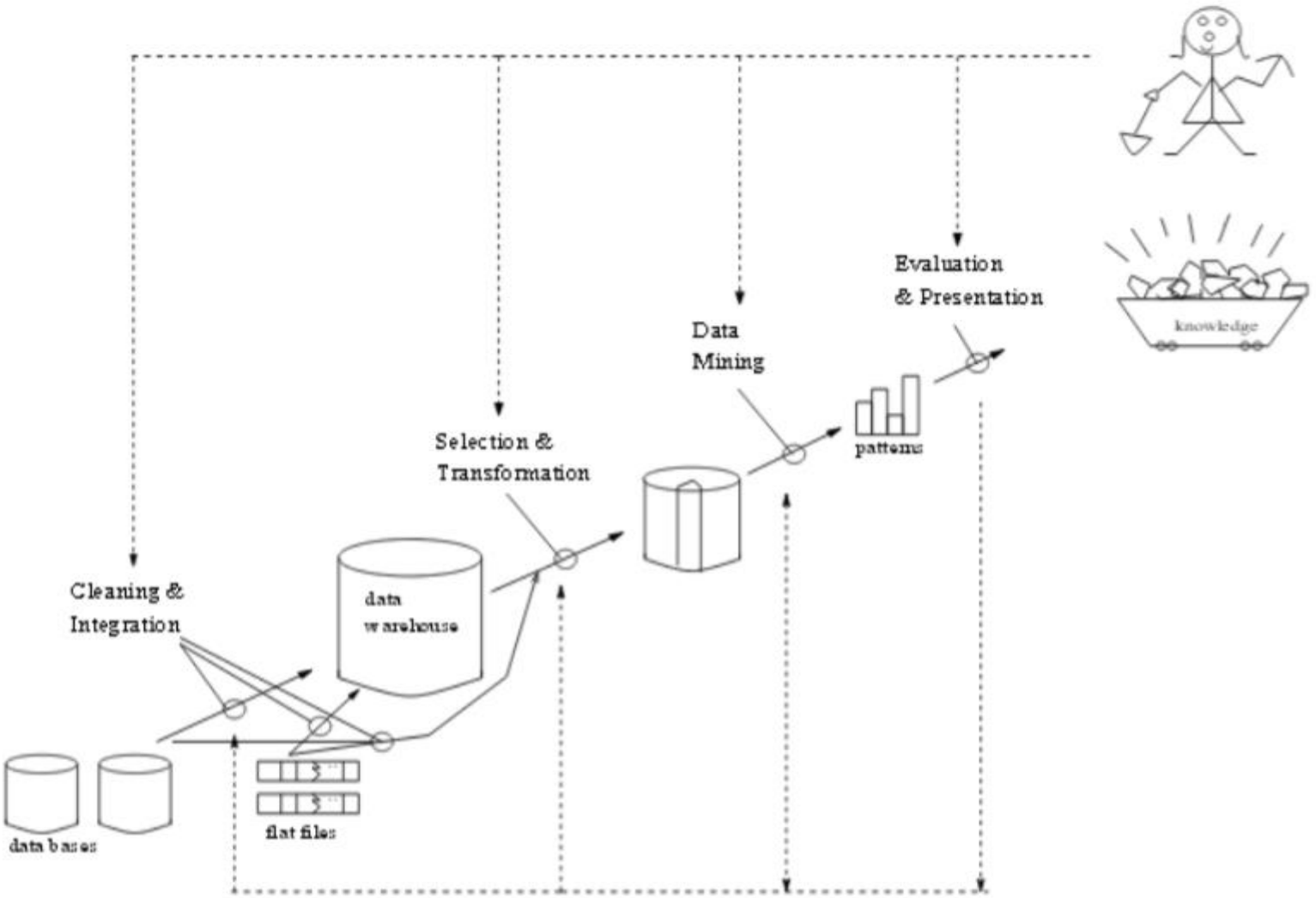
KNOWLEDGE DISCOVERY IN DATABASES (KDD)

- Knowledge discovery in databases (KDD) is the process of discovering useful knowledge from a collection of data.
- This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results.

Here is the list of steps involved in the knowledge discovery process −

- **Data Cleaning** − In this step, the noise and inconsistent data is removed.
- **Data Integration** − In this step, multiple data sources are combined.
- **Data Selection** − In this step, data relevant to the analysis task are retrieved from the database.
- **Data Transformation** − In this step, data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.
- **Data Mining** − In this step, intelligent methods are applied in order to extract data patterns.
- **Pattern Evaluation** − In this step, data patterns are evaluated.
- **Knowledge Presentation** − In this step, knowledge is represented.

KNOWLEDGE DISCOVERY IN DATABASES (KDD)



ARCHITECTURE OF DATA MINING SYSTEM

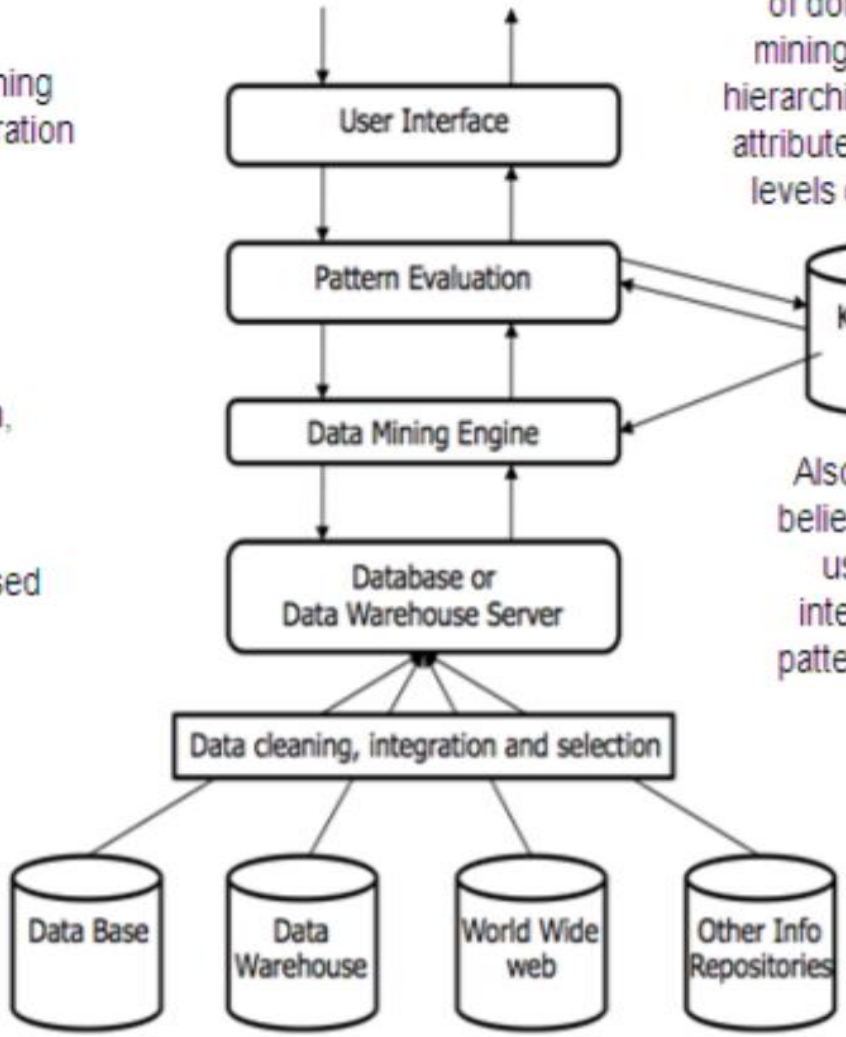
Communicates between users and data mining system. Visualizes results or perform exploration on data and schemas.

Tests for interestingness of a pattern

Performs functionalities like characterization, association, classification, prediction etc.

Is responsible for fetching relevant data based on user request

This is usually the source of data. The data may require cleaning and integration.



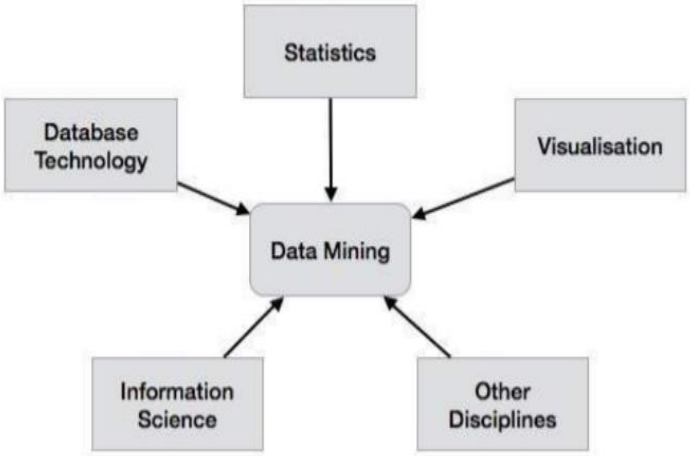
This is the information of domain we are mining like concept hierarchies, to organize attributes onto various levels of abstraction

Also contains user beliefs, which can be used to access interestingness of pattern or thresholds

DATA MINING SYSTEM CLASSIFICATION

A data mining system can be classified according to the following criteria −

- Database Technology
- Statistics
- Machine Learning
- Information Science
- Visualization
- Other Disciplines



Multiple Choice Question

1. Record cannot be updated in _____.

A. OLTP
B. files
C. RDBMS
D. data warehouse

2.. The source of all data warehouse data is the_____.

a) operational environment.
b) informal environment.
c) formal environment.
d) technology environment.

3. Data warehouse contains_ _____ data that is never found in the operational environment.

a) normalized
b) informational
c) summary
d) denormalized

4. The modern CASE tools belong to _____ category.

a) analysis.
b) Development
c) Coding
d) Delivery

5. Bill Inmon has estimated _____ of the time required to build a data warehouse, is consumed in the conversion process.

a) 10 percent.
b) 20 percent.
c) 40 percent
d) 80 percent.

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm

- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf
  DATA MINING BOOK WRITTEN BY Micheline Kamber
- https://www.javatpoint.com/three-tier-data-warehouse-architecture