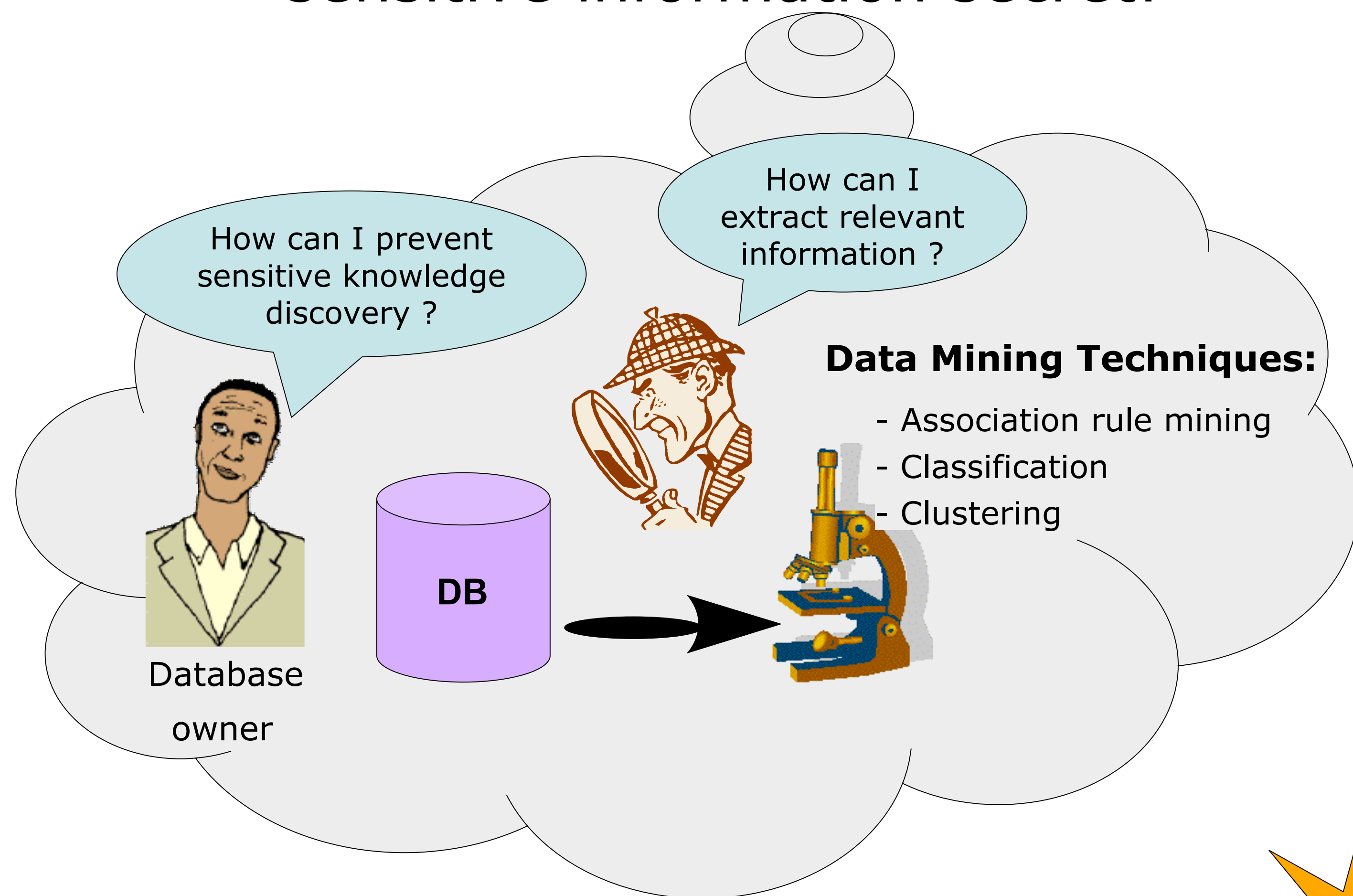


E. Bertino, I. Nai Fovino and L. Parasiliti Provenza

1 Problem :

Extract knowledge from large database while maintaining sensitive information secret!



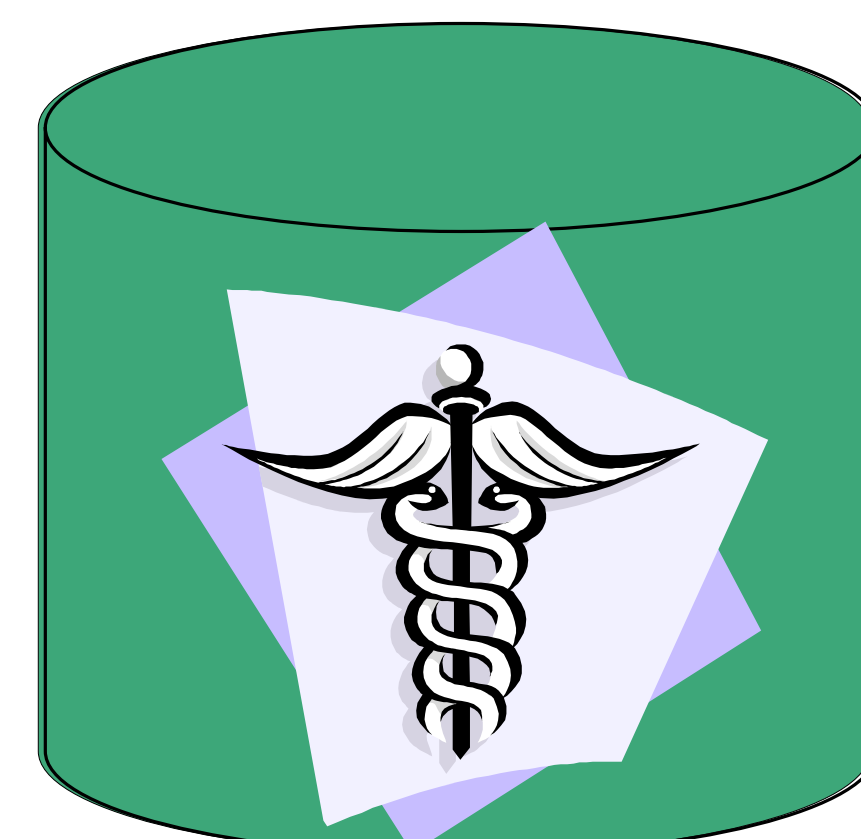
2 What is Sensitive ?

Basket Market DataBase



Sensitive information example:
• If people buy A then people buy also B

Healthcare DataBase



Sensitive information example:
• Specific information about patient pathologies

3 CODMINE Project



- funded by the European Union under the IST Program

Partners:

- University of Milan
- Sabancy University
- Computer Technology Institute

Objectives:

- State-of-the-art
- Development of new algorithms
- Identification of an evaluation framework

4 Classification hierarchy of privacy preserving algorithms

based on the following dimensions:

- **data distribution**
 - centralized databases
 - distributed data
 - horizontally partitioning
 - vertical partitioning
- **data modification** (perturbation, blocking aggregation, swapping and sampling)
- **data mining algorithm**
- **raw or aggregated data hiding**
- **privacy preservation**
 - heuristic-based techniques
 - cryptography-based techniques
 - reconstruction based techniques

5 Proposed algorithms

8 privacy preserving algorithms

- association rule hiding by data perturbation based on a heuristics;
- conceived for centralized databases.

6 Evaluation Criteria

- **Efficiency:** a measure of time and space requirements
- **Scalability:** time trend with data dimension increasing
- **Data quality:** a measure of data degradation after hiding process
- **Hiding failure:** rate of discovered sensitive data
- **Privacy level:** confidence in extracting hidden sensitive data

7 Future Work

- Development and evaluation of new hiding algorithms conceived for classification and clustering techniques
- Investigation of new evaluation criteria

References

- V. S. Verykios, E. Bertino, A. K. Elmagarmid, Y. Saygin and E. Dasseni. **Association Rule Hiding**. IEEE Transactions on Knowledge and Data Engineering (2003).
- V. S. Verykios, E. Bertino, I. Nai Fovino, L. Parasiliti Provenza, Y. Saygin and Y. Theodoridis. **State-of-the-art in Privacy Preserving Data Mining**. To Appear in SIGMOD RECORD.
- E. Bertino, I. Nai Fovino, L. Parasiliti Provenza. **Evaluation of privacy preserving data mining algorithms**. Technical Report CODMINE Project 2003