Data Warehousing, Data Mining, and Business Intelligence

Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.im.tku.edu.tw/~myday/
2011-02-22
Syllabus

1  100/02/15  Introduction to Data Warehousing
2  100/02/22  Data Warehousing, Data Mining, and Business Intelligence
3  100/03/01  Data Preprocessing: Integration and the ETL process
4  100/03/08  Data Warehouse and OLAP Technology
5  100/03/15  Data Cube Computation and Data Generation
6  100/03/22  Association Analysis
7  100/03/29  Classification and Prediction
8  100/04/05  (放假一天) (民族掃墓節)
9  100/04/12  Cluster Analysis
10 100/04/19  Mid Term Exam (期中考試週)
11 100/04/26  Sequence Data Mining
12 100/05/03  Social Network Analysis and Link Mining
13 100/05/10  Text Mining and Web Mining
14 100/05/17  Project Presentation
15 100/05/24  Final Exam (畢業班考試)
Knowledge Discovery (KDD) Process

* Data Warehouse: fundamental process for Data Mining and Business Intelligence
* Data mining: core of knowledge discovery process

Data Warehouse: fundamental process for Data Mining and Business Intelligence

Data mining: core of knowledge discovery process

Source: Han & Kamber (2006)
Data Warehouse

Data Mining and Business Intelligence

Increasing potential to support business decisions

Decision Making

Data Presentation

Visualization Techniques

Data Mining

Information Discovery

Data Exploration

Statistical Summary, Querying, and Reporting

Data Preprocessing/Integration, Data Warehouses

Data Sources

Paper, Files, Web documents, Scientific experiments, Database Systems

Source: Han & Kamber (2006)
Evolution of Database Technology

Data Collection and Database Creation
(1960s and earlier)
• Primitive file processing

Database Management Systems
(1970s–early 1980s)
• Hierarchical and network database systems
• Relational database systems
• Query languages: SQL, etc.
• Transactions, concurrency control and recovery
• On-line transaction processing (OLTP)

Advanced Database Systems
(mid-1980s–present)
• Advanced data models: extended
  relational, object-relational, etc.
• Advanced applications: spatial, temporal,
  multimedia, active, stream and sensor,
  scientific and engineering, knowledge-based

Advanced Data Analysis:
Data Warehousing and Data Mining
(late 1980s–present)
• Data warehouse and OLAP
• Data mining and knowledge discovery:
  generalization, classification, association,
  clustering,
• Advanced data mining applications:
  stream data mining, bio-data mining,
  time-series analysis, text mining,
  Web mining, intrusion detection, etc.

Web-based databases
(1990s–present)
• XML-based database systems
• Integration with information retrieval
• Data and information integration

New Generation of Integrated Data and Information Systems
(present–future)

Source: Han & Kamber (2006)
Evolution of Database Technology

• 1960s:
  – Data collection, database creation, IMS and network DBMS

• 1970s:
  – Relational data model, relational DBMS implementation

• 1980s:
  – RDBMS, advanced data models (extended-relational, OO, deductive, etc.)
  – Application-oriented DBMS (spatial, scientific, engineering, etc.)

• 1990s:
  – Data mining, data warehousing, multimedia databases, and Web databases

• 2000s
  – Stream data management and mining
  – Data mining and its applications
  – Web technology (XML, data integration) and global information systems

Source: Han & Kamber (2006)

- IBM Watson: Final Jeopardy! and the Future of Watson,  
  http://www.youtube.com/watch?v=Il-M7O_bRNg
- Smartest Machine On Earth (2011) 1/4,  
  http://www.youtube.com/watch?v=qIDLd1HUjxY
- Smartest Machine On Earth (2011) 2/4,  
  http://www.youtube.com/watch?v=gg656SKnVQM
- Smartest Machine On Earth (2011) 3/4,  
  http://www.youtube.com/watch?v=hZ7Hsob-h_Q
- Smartest Machine On Earth (2011) 4/4,  
  http://www.youtube.com/watch?v=ozQG_jIB8SE
References

• Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Second Edition, 2006, Elsevier
• Lucene, http://en.wikipedia.org/wiki/Lucene