

Multimedia Data Mining

Jelena Tešić

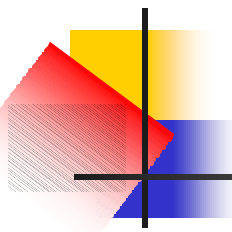
Advisor: B.S. Manjunath

Vision Research Laboratory

Department of Electrical and Computer Engineering

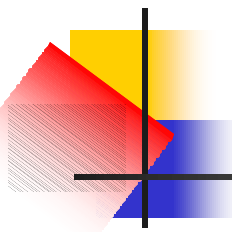
University of California, Santa Barbara





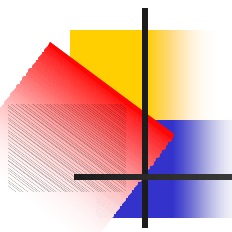
Data Mining

- Data Mining definition:
 - A class of database applications that look for hidden patterns in a group of data.
 - Finding rules of the game knowing the moves of the game
- Unifying framework for data representation and problem solving in order to learn and discover from large amounts of different types of data.



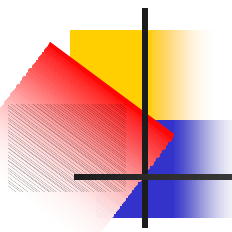
Multimedia Data Mining

- Multimedia data types
 - any type of information medium that can be represented, processed, stored and transmitted over network in digital form
 - Multi-lingual text, numeric, images, video, audio, graphical, temporal, relational, and categorical data.
- Relation with conventional data mining term



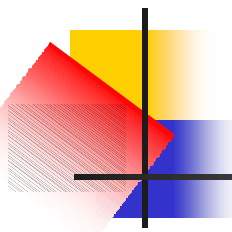
Definitions

- Subfield of data mining that deals with an extraction of implicit knowledge, multimedia data relationships, or other patterns not explicitly stored in multimedia databases
- Influence on related interdisciplinary fields
 - Databases – extension of the KDD (rule patterns)
 - Information systems – multimedia information analysis and retrieval – content-based image and video search and efficient storage organization



Case-base reasoning

- Case representations
 - Structured (KDD applications)
 - Object-oriented
 - Relational attribute-value case
 - Unstructured (multimedia)
 - Limited expressive power
 - Collection of case descriptors
 - Links – connect information within case



Case representation

- Hierarchy of concepts, represented by different views
- Domain decomposition
 - Complex case is represented as multiple cases
- Hierarchy structure supports human reasoning
- Automated process
 - Structured representation layer
 - Vector of case attributes
 - Identify attributes

Case Library

Layer_n

Case

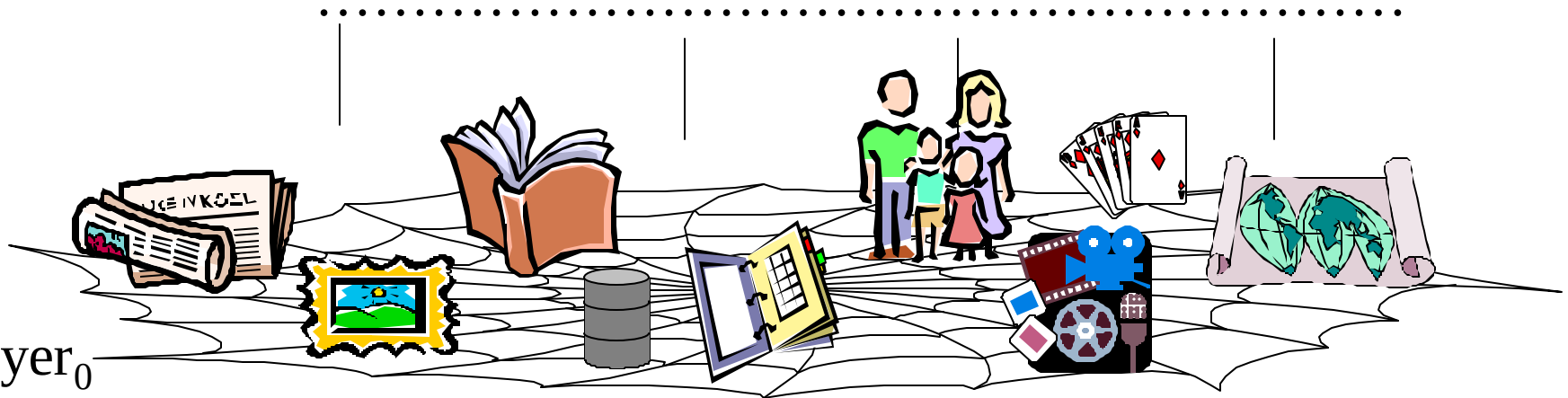
Case

Case

Case

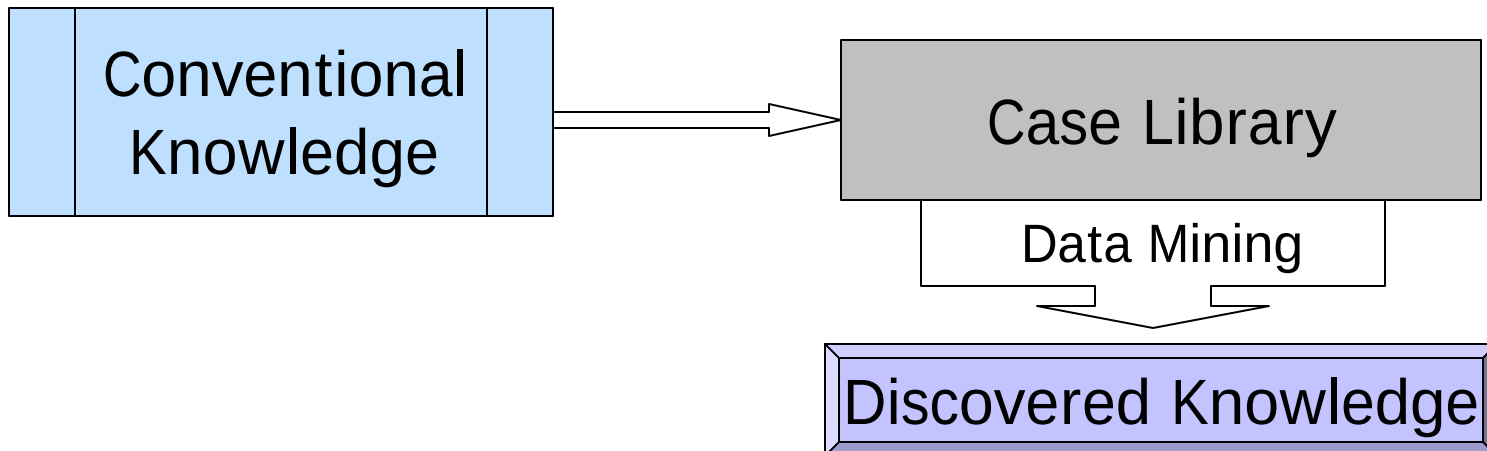
Layer_{n-1}

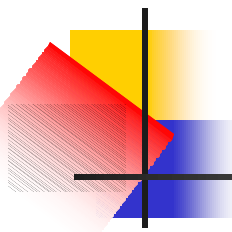
Layer₀



Knowledge Discovery in Multimedia Databases

- Find patterns in primarily unstructured data
- Machine learning where a case library replaces the training set

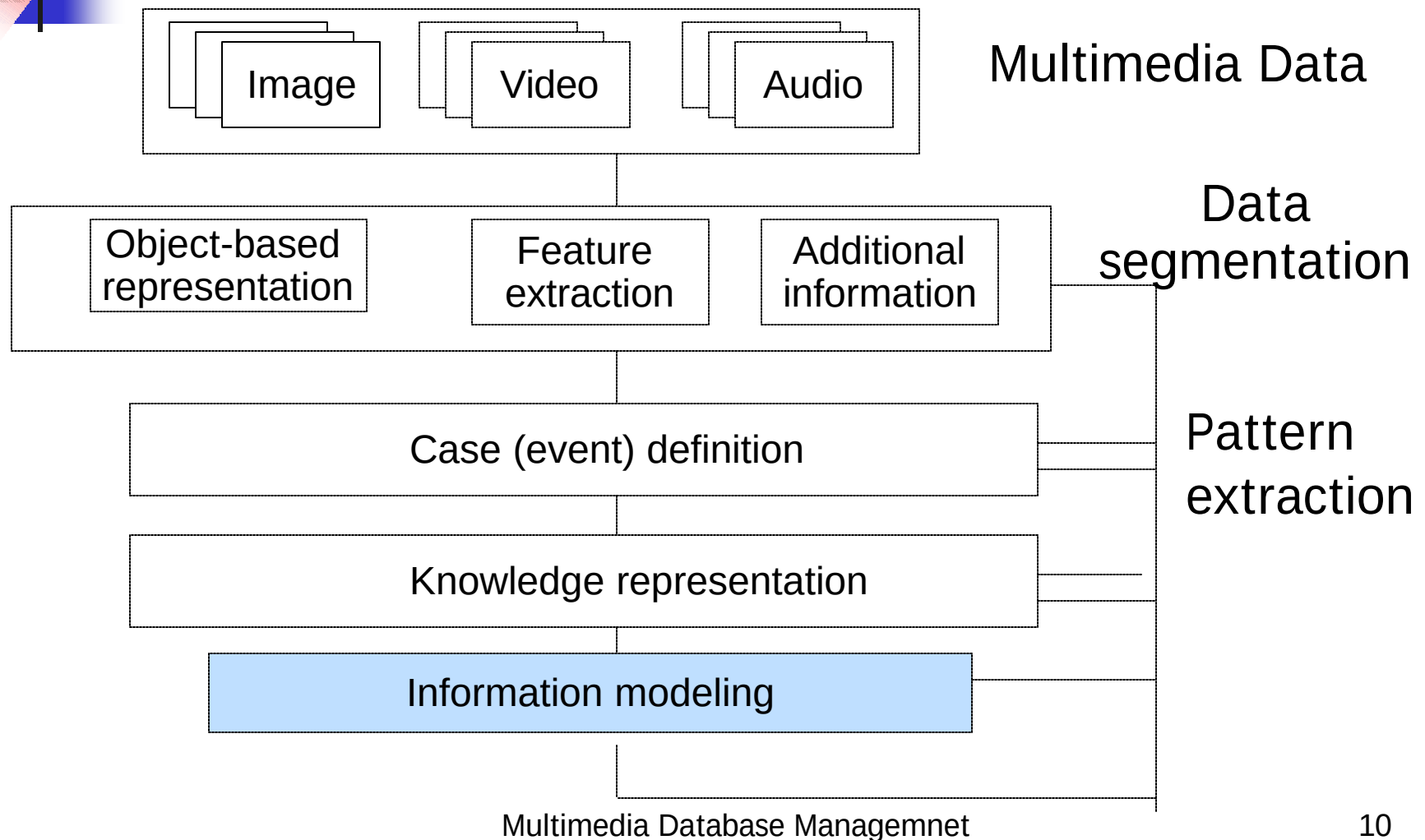


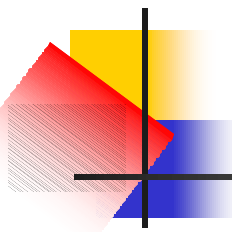


Information model

- Data segmentation
 - Multimedia data are divided into logical interconnected segments (objects)
- Pattern extraction
 - Mining and analysis procedures should reveal some relations between objects on the different level
- Knowledge representation
 - Incorporated linked patterns
- Information model – dynamic structure

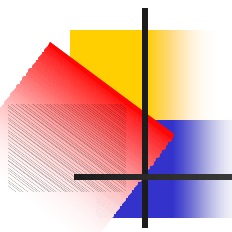
Multimedia Mining Hierarchy





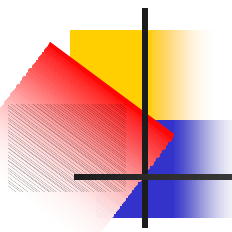
Importance of Case-base reasoning

- Finding patterns based on the specific interest
- Previous experience
 - Assist with indexing and adapting cases to improve retrieval
 - Indication when the adaptation lies outside some reasonable experience
 - Dynamic thematic paths in the hierarchy can assist with navigation in the retrieved cases
- Learning loop of the case-based reasoning



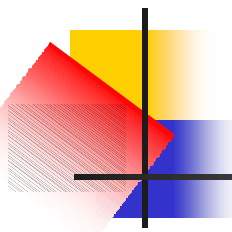
Advantages

- Generation of indexing schemes, based on
 - the related terms to regularities discovered in other media types (semantic extraction)
 - Structural patterns discovered in multimedia (graph indexing)
 - One case library and its dynamic nature
- Retrieval – flexibility in formulating queries
- Adaptation of the new case description based on the user's feedback



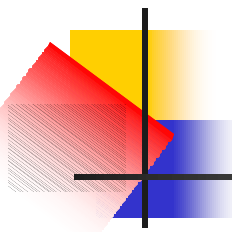
Advantages – cont'd

- Case-based mechanism provides incorporation and management of the discovered knowledge
- Multimedia data mining can improve the case-based system
 - Discover of unknown patterns
- Modular approach to the case-base reasoning multimedia data mining model



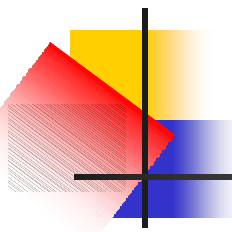
Modular approach

- <http://www.cartogra.com>
- Developer
 - Implementation of any data segmentation and data mining method
 - Adaptation of the stored knowledge
- User
 - Online processing (photo collection)
 - Automatic classification
 - Real time complex query response
 - Feedback



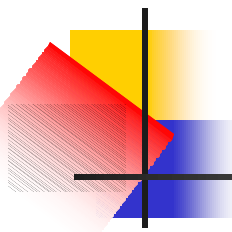
System implementation

- Pattern recognition for larger image databases (Toshiba)
- Content-based retrieval
 - Relationship among features
 - User's feedback (feature weights)
- MultiMedia Miner (Han, SFU, CA)
 - System prototype



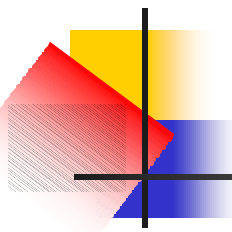
MultiMedia Miner

- Multimedia Data Cube
 - Image Excavator (Extraction of images)
 - Preprocessor - Feature extractor
 - User interface
 - Search engine
- Multimedia Miner
 - characterizer, comparator
 - classifier, associator



Related workshops in 2000

- Workshop on Multimedia Data Mining, Sixth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 20-23, Boston, MA
- Workshop on Mining Scientific Datasets, AHPDR Center, July 20-21, Minneapolis, MN
- Workshop on Data Mining in the Internet Age, IBM Almaden, May 1-2, San Jose, CA



Conclusion

- Multimedia data mining
 - New methodologies
 - Influence on the related fields
- <http://vision.ece.ucsb.edu/~jelena/research/>
- http://www.cs.ualberta.ca/~zaiane/mdm_kdd2000/
- <http://db.cs.sfu.ca>

The End
Thank you