Paper by W. F. Cody
J. T. Kreulen
V. Krishna
W. S. Spangler
Presentation by Dylan Chi
Discussion by Debojit Dhar

THE INTEGRATION OF BUSINESS INTELLIGENCE AND KNOWLEDGE MANAGEMENT

OUTLINE

- Business Intelligence
- Knowledge Management
- BIKM
- eClassifier
- Integrated BIKM Tools

BUSINESS INTELLIGENCE

*"**Business intelligence (BI)** refers to skills, technologies, applications and practices used to help a business acquire a better understanding of its commercial context."*

*"**Business intelligence** may also refer to the collected information itself."*

--Wikipedia

BUSINESS INTELLIGENCE

- Business intelligence technology has coalesced around the use of two technologies
  - data warehousing
  - on-line analytical processing (OLAP).

DATA WAREHOUSING

Data warehousing is a systematic approach to collecting relevant business data into a single repository, where it is organized and validated so that it can be analyzed and presented in a form that is useful for business decision-making.

DATA WAREHOUSING

- The various sources for the relevant business data are referred to as the **operational data stores (ODS)**.
- The data are **extracted**, **transformed**, and **loaded** (**ETL**) from the ODS systems into a data mart.

OLAP

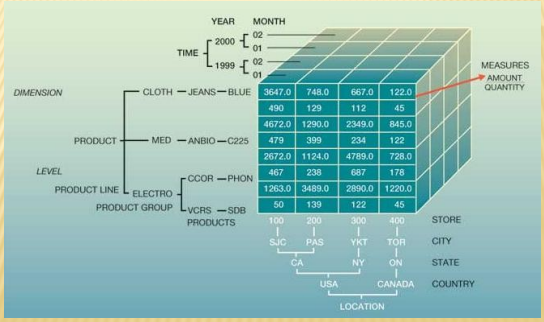
*"**Online analytical processing**, or **OLAP**, is an approach to quickly answer multi-dimensional analytical queries."*

--Wikipedia

OLAP

- In the data mart, the data are modeled as an OLAP cube (multidimensional model)
- Multidimensional model supports flexible drill-down and roll-up analyses

OLAP CUBE



OUTLINE

- Business Intelligence
- Knowledge Management
- BIKM
- eClassifier
- Integrated BIKM Tools

KNOWLEDGE MANAGEMENT

*"**Knowledge Management** (**KM**) comprises a range of practices used in an organization to identify, create, represent, distribute and enable adoption of insights and experiences. Such insights and experiences comprise knowledge, either embodied in individuals or embedded in organizational processes or practice."*

--Wikipedia

KNOWLEDGE MANAGEMENT

- In this context, it is used for the management and analysis of unstructured information, particularly text documents.
- Textual information sources
  - Business documents, e-mail, news and press articles, technical journals, patents, conference proceedings, business contracts, government reports, regulatory filings, discussion groups, problem report databases, sales and support notes, web.

DISCUSSION

- Does the authors' definition of business intelligence agree with yours? Why or why not?
- What business intelligence applications can you think of that aren't mentioned in the paper?

OUTLINE

- Business Intelligence
- Knowledge Management
- BIKM
- eClassifier
- Integrated BIKM Tools

BIKM

- The authors believe that over time techniques from both BI and KM will blend
- New techniques will seamlessly span the analysis of both data and text

BIKM PROBLEMS

- Understanding sales effectiveness
  - Products, sales representatives, customers
  - Sales techniques
- Improving support and warranty analysis
  - Customer complaints
- Relating CRM to profitability
  - 'hidden' cost
  - Complete picture

ENVIRONMENTAL ISSUES

- Text information sits inside the same database
- Textual information is in systems distinct from the ODS systems
- The sources of text to relate to a business data analysis are not known

EXAMPLE

- A business analyst explore a revenue cube and detect a downward movement in revenues for a software product in some part of the United States.
- The data cube shows the phenomenon but does not provide any explanation for it

EXAMPLE

- To understand the phenomenon, some text sources could be used to extract valuable information
  - Enterprise-specific information
    - Service call logs about the product
    - Competitive intelligence reports
  - Purchased text information
  - Public documents in Web forms
    - Discussions about products

DISCUSSION

- Do you think integrating BI and KM to be a good idea?
- Do you think the ideas in the paper made/did not make it to the mainstream BI tools? Have you come across tools that use the BIKM concept?

OUTLINE

- Business Intelligence
- Knowledge Management
- BIKM
- eClassifier
- Integrated BIKM Tools

ECLASSIFIER

- **eClassifier** is an application that can quickly analyze a large collection of documents and utilize multiple algorithms, visualizations, and metrics to create and to maintain a taxonomy.
- It is very difficult to automatically produce a satisfactory taxonomy for a diverse set of users without allowing human intervention.

DOCUMENT REPRESENTATION

- Feature space of terms and phrases
  - The feature space is obtained by counting the occurrence of terms and phrases in each document
  - Stop-word lists
    - Synonym list, stock phrase list, 'include word' list
- Vector of weighted frequencies
  - Dictionary tool

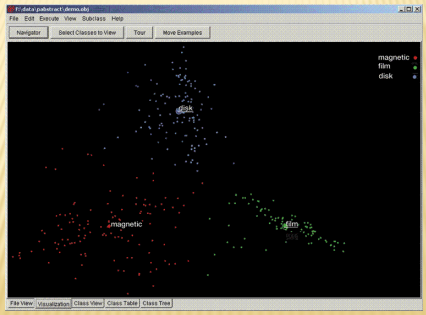
TAXONOMY GENERATION

- Automatically create an initial categorization or taxonomy
  - *k*-means algorithm
- Interactive, query-based clustering
  - Seeds categories based on a set of keywords
  - Tests out the queries
  - Refines the clusters based on the observed results

TAXONOMY EVALUATION

- Once we have an initial taxonomy of the documents, eClassifier provides the means to understand and to evaluate it.
- Category label is generated using a term-coverage algorithm that identifies dominant terms in the feature space.
- Metrics
  - Size, cohesion, distinctness

TAXONOMY VISUALIZATION



CLASSIFICATION

- Assign additional documents to the taxonomy as they become available
- eClassifier creates a batch classifier to process the additional documents
  - Nearest centroid
  - Native Bayes multivariate
  - Native Bayes multinomial
  - Decision tree

ANALYSIS AND REPORTING

- FAQ analysis
- Discovery of correlations
- Chi-squared test
- Continuous variables
- Using a generated taxonomy to compare document collections
- …

DISCUSSION

- The main tasks of eClassifier can be represented as:
  - Taxonomy generation
  - Taxonomy and category evaluation
  - Taxonomy visualization
  - Classification
  - Analysis and reporting

  Which of these do you think is most important and why?

OUTLINE

- Business Intelligence
- Knowledge Management
- BIKM
- eClassifier
- Integrated BIKM Tools

INTEGRATION PARADIGM

- Text is ultimately associated with business data records to enhance the understanding of the data
- We might strive to achieve a tighter integration of the text information with the associated data
  - Using an OLAP multidimensional data model as the integrating mechanism

INTEGRATING TEXT INFORMATION

- Find attributes in the documents that can be used to link them to the data, or find attributes in the documents that can be used as additional dimensions to deepen the understanding of the data
- Compute quantitative values from the documents

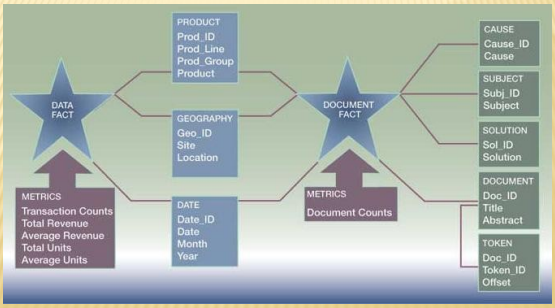
INTEGRATED BIKM TOOLS

- Apply the OLAP data model to text documents, creating a document warehouse
- Allow users to explore data cubes with a star schema and consists of a report view and navigational controls

DOCUMENT WAREHOUSING

- The fact table granularity is a document
- The dimension tables hold the attributes of the document

SHARED DIMENSION DATA MODEL



SHARED DIMENSIONS

- We use star schemas to organize and analyze both data and document cubes
- Providing a mechanism to link them will allow deeper analysis and thereby provide greater value
- The key to achieving it is to directly link the data to the documents through shared dimensions

DYNAMIC DIMENSIONS

- The new taxonomy can be made available to the document warehouse by creating a corresponding dimension table to represent the taxonomy and then populating an added column in the fact table, associating all known document with the newly published dimension

DISCUSSION

- Do you think it is possible to create a consistent taxonomy for documents using the concepts detailed in the paper? What changes would you suggest to come up with a more useful classification?
- Is it a good idea to group documents under a single hierarchy or class?

Thank you