Text Mining

What is Text Mining

1. It is the analysis of data contained in natural language text
2. It works by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques
3. In text mining, the goal is to discover heretofore unknown information, something that no one yet knows and so could not have yet written down.

Process of Text Mining
1. Definition of mining context and concepts
2. Data collection.
3. Dictionary construction (End Note 1)
4. Data analysis

Difference with Data Mining
In Text Mining, patterns are extracted from natural language text rather than databases.

Example
Hewlett-Packard (HP)
Hewlett-Packard wants to make better decisions based on what customers want and expect in the products and services they need. They use text-mining tools to analyze e-mail, customer surveys, warranty claims forms, and feedback. Text mining tools enable them to (1) uncover underlying themes contained in large document collections, (2) automatically group documents into topical clusters, (3) classify documents into predefined categories, (4) integrate text data with structured data to enrich predictive modeling endeavors, and (5) base business decisions on insightful customer information.

Reference URLs:
http://searchbusinessanalytics.techtarget.com/definition/text-mining
http://www.cs.sunysb.edu/~cse634/presentations/TextMining.pdf
http://people.ischool.berkeley.edu/~hearst/text-mining.html

Google Search Keywords
examples of text mining, text mining examples, examples of text mining in business, short examples of text mining in business

Endnote:
1. (Dictionary construction: As we explained above, researchers must construct a dictionary to associate search terms with specific concepts". URL: http://www.provalisresearch.com/Documents/CHRMQ2.pdf, Page: 348)

Submitted by
S. M. Benzir Ahmed
Id: 00-00000-0
Graduate Student, MBA (Regular), AIUB, Dhaka
Course: Data Warehouse & Data Mining
Date of Submission: 3 August 2011