

showcase of a life



www.shaon84.com



Presentation On
Text Mining
For The
Hotel Industry

Presentation On
Text Mining
For The
Hotel Industry

Welcome

To The Presentation



Presentation On
Text Mining
For The
Hotel Industry

Presentation By

S. M. Benzir Ahmed

Id: 00-00000-0

Graduate Student (MBA)

American International University - Bangladesh



Presentation On
Text Mining
For The
Hotel Industry

Prepared For

Mohammad Saidur Rahman

Assistant Professor

Department of Computer Science & Engineering

American International University - Bangladesh



Presentation On
Text Mining
For The
Hotel Industry

Course

Data Warehouse & Data Mining
An MIS Major Course



Presentation On
Text Mining
For The
Hotel Industry

Main Article By

Prof. Dr. Kin-nam Lau
Prof. Dr. Kam-hon Lee
Ms. Ying Ho



Presentation On
Text Mining
For The
Hotel Industry

Article Publication Detail

Cornell Hotel and Restaurant Administration Quarterly
Cornell University
Volume 46, Number 3 344-362, August 2005
Page: 344-362



Introduction

- § Business intelligence plays an important role
- § Proper competitor and customer intelligence enhances hotel effectiveness and customer satisfaction
- § Hospitality practitioners are overloaded by data
- § Many business information exists in the form of unstructured or semi-structured text documents
- § Traditional text information processing requires substantial investment of money, time, and human resources
- § It is hard to combine qualitative text data with quantitative numeric data
- § Need a method to accurately extract business intelligence from large text collections and integrate the fragmented information into business intelligence databases

What This Article Covers

- § Proposes text mining as a means of information management
- § That can analyze the huge textual information found in a hotel's internal databases & external sources
- § Based on the demonstration study, this article discusses potential uses
- § Technological limitations of text mining
- § Addresses the implementation cost
- § Possible future technological advancement



Text Mining: Concept

- § Proposes text mining as a means of information management
- § That can analyze the huge textual information found in a hotel's internal databases & external sources
- § Based on the demonstration study, this article discusses potential uses
- § Technological limitations of text mining
- § Addresses the implementation cost
- § Possible future technological advancement

Text Mining: Concept

- § Text mining explores data in text files
- § Establish valuable patterns and rules that indicate trends and significant features about specific topics
- § Works with an unstructured or semi-structured collection of text documents
- § Example: corporate documents, Web pages, newsgroup postings
- § Process starts with a keyword search in text collections
- § It is necessary for researchers to construct a dictionary that acts as the knowledge base to associate keywords with specific concepts.
- § Text Mining Data for Hotailers



Text Mining: Data for Hotailers

- § Environmental scanning of customer intelligence by analyzing customer newsgroups, online bulletin boards, and online customer surveys;
- § Acquiring customer intelligence by analyzing personal home pages, customer comment cards, and qualitative survey data; and
- § Improving efficiency of internal knowledge management by analyzing e-mail, patent databases, and corporate documents.

Data Collection: Traditional Research VS Text Mining

Similarities

- §The text-mining concept is similar to traditional survey research
- §The process of constructing a keyword dictionary is similar to constructing a questionnaire
- §In the process of text search, the search engine is the “interviewer,” while the individual text document is the “interviewee.”



Data Collection: Traditional Research VS Text Mining

Difference

- § traditional research seeks to understand the target population
- § objective of text mining is to study the whole population instead of just a sample
- § survey questions in marketing research are developed according to the researchers' specific interests
- § data in the text-mining approach are self-revealed according to the owners' preferences

Online Text Mining Basic

- § searching through the volumes of material available on the Internet
- § Internet generates and stores considerable amounts of business information in online databases, such as company Web sites, customer newsgroups, and online focus groups



Online Text Mining Example

- § knowledge about potential customers (e.g., gender, age, marital status, interests, and hobbies) is available on personal home pages
- § Discussions in newsgroups and online bulletin boards may serve as abundant sources of market intelligence (e.g., consumer preferences, evaluation of existing services, and customer complaints)
- § Key “firmographics” (e.g., number of hotel rooms, price plans, services, and facilities) can be extracted from hotel Web sites

Online Text Mining

Step of Text Mining

1. Definition of mining context and concepts
 - identify the types of information being sought
2. Data collection
 - establish which text documents will be analyzed
3. Dictionary construction
 - construct a dictionary to associate search terms with specific concepts
4. Data analysis
 - § translate unorganized text into meaningful figures and indexes
 - § qualitative text data can be combined with quantitative data for more comprehensive business analyses
 - § interpret results for managerial decision making



Demonstrating the Potential of Text Mining

Studies

- § Study 1: A Hotel Profile (Invariant Information)
- § Study 2: Room Prices (Variant Information)
- § Study 3: Travel-Related Newsgroups (Customer Intelligence)



Presentation On
Text Mining
For The
Hotel Industry

Demonstrating the Potential of Text Mining

Note That

- § The illustrations do not imply that text-mining technology is readily applicable in the hotel industry at the moment
- § The purpose is to show what might be possible



Presentation On
Text Mining
For The
Hotel Industry

Competitive Intelligence

About Hotel Data

- § August 2002, they gathered Web pages of seventy-four member hotels of the Hong Kong Hotels Association
- § After preliminary screening, they removed seven hotels from further analysis
- § A total of sixty-seven hotels were retained for text-mining analyses



Competitive Intelligence

Types of Competitive Intelligence

Competitive intelligence in the hotel industry can be classified as

- § Invariant competitive intelligence
- § variant competitive intelligence

Competitive Intelligence

Invariant competitive intelligence

§ Remains unchanged most of the time and thus may not require close monitoring for day-to-day changes

§ **Examples** include such attributes of a hotel's profile as room type, food, in-room amenities, business services, beauty services, health and sports facilities, other services, affiliated hotels, and contact information

Competitive Intelligence

Variant competitive intelligence

- § Frequently updated
- § Changes can be important indications of competitive moves and should be closely monitored.
 - § **Examples** include room price and promotion packages

Competitive Intelligence

Measuring Accuracy

- § text-mining accuracy should be defined as how well the text mining result is a true representation of the text collection
- § They compared the computer's output with the human analyst's results and calculated the hit rate for each concept
- § define hit rate percentage as the number of truly represented cases compared to all analyzed cases, as follows:

$$\text{Hit Rate (\%)} = \frac{\text{Correctly Classified Cases}}{\text{Total Number of Cases}} \times 100\%.$$

Study 1: A Hotel Profile (Invariant Information)

Findings

- § This study aims at constructing a database that contains profile information of Hong Kong hotels (e.g., services and facilities, location, contact method)
- § hotel profile database useful to hoteliers in identifying opportunities and threats in current and prospective markets
- § The hotel profile dictionary includes 18,340 search terms
- § The problem of missing values is most noticeable for beauty services and affiliated hotels, where missing values exceed 37 percent
- § All concepts except in-room amenities and contact information have hit rates above 90 percent

Study 1: A Hotel Profile (Invariant Information)

Findings (Cont...)

- § Results for those two concepts are less than satisfactory
- § results suggest that the text-mining technique can generate reasonable accuracy for a hotel profile analysis
- § **Two factors** affect the levels of missing values and hit rates
- § First, the dictionary could be made more exhaustive with the addition of more keywords
- § Second, technical limitations exist for context-based query and Web-table analysis
- § For example, our text-mining tool cannot recognize sentence breaks except by conventional punctuation, namely, periods, question marks, or exclamation marks
- § hotel Web sites is displayed in tables and lists that do not include punctuation, some data are overlooked, bringing in inaccurate results

Study 2: Room Prices (Variant Information)

Findings

- § hotel room prices may change continually For simplification, we adopted a static view of a hotel's pricing schedule

- § Before actual analysis, we processed Web pages by
- § 1. reformatting the source file of a home page into a tag-free single-line sentence,
- § 2. removing Chinese characters, and
- § 3. converting prices given in U.S. dollars into HK dollars.

Study 2: Room Prices (Variant Information)

Findings (Cont...)

- § constructed 158 search terms for the price study
- § It filtered irrelevant prices and numbers such as restaurant meal prices and international direct dialing (IDD) cost
- § excluded room prices with date or promotion keywords
- § We then matched the room-type keyword with the nearest number following it
- § Proximity between a keyword and its corresponding number could not exceed ten words

Study 2: Room Prices (Variant Information)

Findings (Cont...)

- § Results reveal that we successfully identified 83 percent of the normal price information provided in hotel Web sites
- § In identifying room types, all room types have hit rates above 92 percent.
- § In identifying price ranges, we achieve hit rates above 90 percent for practically all room types
- § The only exception is harbor-view rooms, which achieved a less satisfactory hit rate of 67 percent.

Study 3: Travel-Related Newsgroups (Customer Intelligence)

Findings

- § Travelers' demographic profiles, preferences, and interests are valuable information for hoteliers
- § Study 3 examined newsgroup postings related to traveling in Europe
- § The objective is to understand the demographic profile, primary interests, and concerns of potential travelers
- § We collected 4,393 newsgroup articles that were posted from March 29 through April 24, 2002.
- § This dictionary includes 6,657 travel-related keywords for the analysis of newsgroup postings.



Study 3: Travel-Related Newsgroups (Customer Intelligence)

Findings (Cont...)

- § After textual information was converted into figures and indexes, we used chi-square analysis to identify possible associations between concepts
- § We explored how gender may affect tourists' travel patterns (e.g., activities and destination choice).
- § We also investigated how travelers' accommodation concerns are associated with their demographic characteristics
- § Newsgroup postings are typically short, concise messages about a single particular topic. As different messages cover different issues, it is reasonable to expect a large missing percentage for each concept.

Study 3: Travel-Related Newsgroups (Customer Intelligence)

Findings (Cont...)

- § Results show that the most commonly identified concepts are destination and gender, which have missing value of less than 41 percent
- § Reservations, activities, ticketing, marital status, and travel with or without children are less often mentioned by newsgroup participants. Missing values of these concepts exceed 90 percent
- § In terms of text-mining accuracy, all concepts except gender attained hit rates above 94 percent
- § Gender has a comparatively less satisfactory hit rate of 83 percent.

Implementation Costs

Major Spending Areas

- § the fixed cost of purchasing a text-mining tool
- § the variable cost of dictionary construction
- § program customization, and maintenance.
- § Variable costs are driven by human labor to build the dictionary and execute and maintain text-mining programs.
- § This labor cost escalates with increasing complexity and execution period of the project

Current Limitations

- § Image files
- § Dynamic Web sites
- § Context-based analysis
- § Region-specific dictionary.

Conclusion

- § It may effectively reduce the use of manual labor in identification, shortage, and analysis of business intelligence.
- § existing text-mining tools are not mature enough to accurately analyze dynamic Web sites and image and animation files.
- § the proposed text-mining approach is a hybrid method that combines efforts of computer programs (profile and price identification) and manual labor (Web page preprocessing).

Thanks for The Support
Any Question Is Welcome...



showcase of a life



www.shaon84.com

