Journal of International Technology and Information Management

Volume 17
Issue 3 *Double Issue 3/4*

Article 2

2008

Identifying Opportunities in Multilingual Business Environments Using Environmental Scanning and Text Mining Techniques

Jody Strong
University of Mississippi

Kaushik Ghosh
University of Mississippi

Sumali Conlon
University of Mississippi

Identifying Opportunities in Multilingual Business Environments Using Environmental Scanning and Text Mining Techniques

Jody Strong
Kaushik Ghosh
Sumali Conlon
University of Mississippi
USA

ABSTRACT

The identification of opportunities for growth can be made easier if comprehensive information relevant to the business environment is available to managers. Such recognition of business opportunities can also help sustain competitive advantage. Information relevant to business environment is usually written and posted in many languages and can be accessed from many sources. The collection of this information is time consuming and labor intensive and techniques such as environmental scanning that are proposed in previous research can facilitate this information search. In this study, we propose a technique to automatically perform tasks using text mining tools that search, translate, and extract information from online documents. Updated information produced by these tools will be current and accessible by all levels of management and facilitate managerial decision making.

INTRODUCTION

As organizations expand their operations in different countries, information concerning the local economy, specific information concerning customers, and competitors must be available to the management to make timely decisions. This helps organizations succeed and achieve competitive advantage. The timely processing and transferring of information to decision makers has been defined in different ways in past research. For this research we will use the term Environmental Scanning (ES).

ES has been described recently by Decker, Wagner, and Scholz (2005) as a means by which managers study relevant business environments. Culnan (1983) described ES as the acquisition of information about events taking place outside the organization which can be used to respond effectively to changes in the environment. Hambrick (1982) described ES as the key step in the process of organizational adaption to the business setting. In summary, ES is acquiring information to allow effective decisions to be made by management.

An organization must decide what information is important and what is not. There are three sources of information that organizations typically access: human, documentary, and physical phenomena (Keegan, 1974). Many executives prefer direct human intelligence. But, with the increasing use of the internet for accessing and dissemination of information, and the need for timely information, the Internet is a quality resource for information. Recently, He and Zhu (2007) looked at corporate blogs to access the information gained from them.

Aguilar (1967) identified the information required by an organization and separated it into several areas (for example, market information, information about competitors etc). Some organizations have developed ES processes, but instead of an overall process, they have developed several processes which include competitive intelligence (CI), knowledge management (KM), and business intelligence (BI). These processes are related to ES in that ES is the action of obtaining information about all aspects of a business. Such processes are usually departmentalized and the information is not always combined to form actionable information in a timely fashion. Specifically many departments control the information for political power within the organization. Most marketing departments focus on competitor information and customer relationship management, while the production and operations department controls the supply chain management system. So in each organization, we must define what processes are developed and what processes need to be developed and combine these separate processes into an overall system to improve the organizations ability to make effective decisions with the information available.

Information relevant to the business environment may be written and posted in many languages. So, the translation of information (in addition to the search and storage of the information) becomes essential. In this paper, we attempt to develop an environmental scanning tool that will perform structured and unstructured scanning of the multilingual business environment, subject to the criteria set by management to deliver timely and relevant information. Timeliness of the information can have an impact on quality of strategic decisions. Thus searching for this information and transferring this knowledge in-time to management is essential.

In the following sections of this paper, we will discuss the motivation for this research, as well as the future research possibilities for ES. In addition, the tools and methods used to extract information automatically are developed. Further, since information on the internet is continually expanding, and varied, and published in many languages, we talk about how MT systems are facilitating the translation of information from one language (source) to another (target), subsequently enhancing the capabilities of ES. Finally, an operational model that stipulates the functions and steps needed for ES are proposed and discussed and conclusions are drawn.

Motivation

Organizations interested in expanding into international markets need comprehensive information about the business environment in those regions to make strategic decisions. To access this information, corporations need to convert local information to the language of interest (which is understood by managers).

As stated earlier, a key factor mentioned in previous literature for a business to succeed is the use of ES (Dollinger, 1984; Daft, Sormunen, & Parks, 1988; Subramanian, Fernandes, & Harper1993; Ngamkroeckjoti & Johri, 2003). Salton (1970) was one of the first to have examined automatic translation of documents from one language to the other. Applying Salton's idea of machine translation in the context of converting business information from one language to the other (target language or language of interest), we illustrate how the combination of ES and MT systems can improve the quality of information (accuracy and timeliness) available to multinational corporations.

Muralidharan (2003) surveyed multinational corporations and found that these firms perform-macro ES, financial ES, and market competitive ES. The first, macro ES, includes societal attitudes toward foreign companies in the foreign country, the general demographic trends in the foreign country, government regulations on foreign investments, trade pacts involving the foreign country, and technology development in the foreign country. The second area, financial ES is composed of inflation rate in the foreign country, the prime lending interest rate in the foreign country, the prices of specific raw materials in the foreign country, and the exchange rate of the foreign country's currency. The last area, market competitive ES included competitor actions in the foreign country, and the market response to the multinational company in the foreign country.

Keeping in mind the above mentioned areas of ES, several ES and CI prototypes have been built by research scholars and commercial software firms. Liu, Turban, and Lee (2000) developed MasterScan for the pulp and paper industry. CI Spider was developed by Chen, Chau, and Zeng (2002). McManus and Snyder (2003) developed EPSS to gather information from the internal sources of an organization. Decker, et al. (2005) used the Information Foraging Theory (IFT) of Pirolli and Card (1999) to develop their system which was capable of performing ES. The Fuld and Company reviewed seventeen ES/CI commercial software packages in their "2006-2007 Intelligence Report." The report provided a two page review and analysis of each of the seventeen ES/CI packages and using metrics (such as ability to conduct meta-searches, filtering of extracted information etc) for comparisons, rated each of the ES/CI software.

In this paper we attempt to investigate ways to improve ES by using available technologies which gather and extract information from the internet and other electronic sources using IE and MT. Using these tools, we search, extract, translate, store, and present the information in the language of interest in a concise format to facilitate strategic decision making.

Further, in this paper, by targeting the area of finance, specifically the prime lending rate in India, and obtaining the data found in articles online, we demonstrate the usefulness of the tools. Finally, using an English to Hindi translator, we show how a MT system helps us to acquire information written in different languages.

Text Mining Tools

Text mining is the process of acquiring information by analyzing and deriving patterns from textual data. Generally, text mining techniques are based on research areas that include information retrieval, data mining, machine learning, statistics, and computational linguistics. Typical text mining tasks consist of text categorization, text clustering, concept/entity extraction, and document summarization. Three of the major text mining techniques are discussed in the following sections.

Information Retrieval (IR)

There are several technologies that are used to retrieve the documents that pertain to the information an organization needs. IR takes input from a web browser and searches the internet

or particular Uniform Resource Locator (URL) address for documents that match the query provided by the user. This search is performed using the HTML, XML and other tags located in the documents that identify the subject of the document.

Information Extraction (IE)

Information Extraction (IE) is a sub-area of Natural Language Processing (NLP) that extracts information from unstructured text documents and produces structured format data. Extracted data can be put in databases or filled in slots in templates (Cardie, 1997; Cowie & Lehnert, 1996). Also, Kalczynski (2005) proposed Temporal Document Retrieval Model (TDRM) which is a knowledge warehouse for storage of extracted documents. We will still use the template model because it does include dates in each template.

In general, IE systems are implemented using two major techniques - the knowledge engineering approach and the automatic training approach (Appelt & Israel, 1999). In knowledge engineering approach, the knowledge engineers who have technical skills gather knowledge from domain experts and build the rules for the systems. This approach can be expensive if the knowledge engineers and the domain experts are not available. The automatic training approach of IE uses large amounts of data to train the systems. If there large amounts of data are not available, this approach might not be appropriate (Appelt & Israel, 1999; Manning & Schutze, 2002).

"Precision" and "recall," refer to the standard measures to assess the accuracy of an IE system (Salton & McGill, 1986). Precision is the accuracy of the IE system compared to what an expert would obtain, while recall is the percentage of information found by the IE system compared to what an expert would find from the same data source. Thus, precision and recall are defined as:

$$\text{Precision:} \quad \frac{\text{Correct entries produced by the system}}{\text{Total number of entries the system produced}}$$

$$\text{Recall:} \quad \frac{\text{Correct entries produced by the system}}{\text{Total number of possible correct entries}}$$

Machine Translation (MT)

Machine translation (MT) is defined as the computer-based technique to automate the process of translating one human/natural language to another. MT generally uses natural language processing (NLP) and usually defines rules for fixed constructions. The original or source text is encoded in a symbolic representation from which the translated text is derived. MT systems are usually classified by their architecture - the overall processing organization, or the abstract arrangement of its various processing modules.

There are several reasons for the growing popularity of MT based systems. As multinational organizations set-up business in foreign markets, they might require large text documents (written in foreign language) to be converted to the language that their employees and/or managers can understand in a short duration. Further, companies who opt to do business over the internet need to attract international customers. Hence, it is necessary that such companies keep

information (e.g.,, company profile and product profile) in a variety of languages. Also, MT systems can substitute for human translators who can be hard to find. Traditional MT systems have been based on direct or transformer architecture engines, and this is still the architecture found in many of the more well-established commercial MT systems.

Machine translators that produce quick and not very accurate translation of the source text to target text are inexpensive while those that provide very reliable translations are sophisticated programs that can cost a considerable amount of money to the buyer (of the MT system). MT systems can be unidirectional, or bi-directional (for example, English to Hindi, as well as Hindi to English).

Machine translation systems based on transformer engines convert input (source) sentences/text into target (output) or desired sentences/text by replacing the source words with their corresponding target language words as specified in bilingual dictionary, and then re-arranging the order of the words to convey the meaning of the target language sentence.

Another relatively new architecture that MT based systems rely on is the linguistic knowledge (LK) architecture. In this, the translation is based on the extensive knowledge the MT system has of both the source and the target languages and of the relationships between analyzed sentences in both languages. Further, the MT system based on the LK architecture requires substantial grammar of both the source and the target language. Also, knowledge of the additional comparative grammar which can be used to relate every source sentence representation to some corresponding target language representation is essential.

SYSTRAN, a widely used MT system which was first installed in the 1970s at the US Air Force is based on the transformer architecture. Some other popular MT systems are EUROTRA and MIMO. MT systems can be error prone in that the translated text might not communicate the same meaning as the original text. Such errors can be eliminated by modifying the information in the lexicon or dictionary. Other forms of error in translating the source to the target text might be due to problems in the MT system's grammar or linguistic processing strategies which cannot be easily resolved without experts' knowledge.

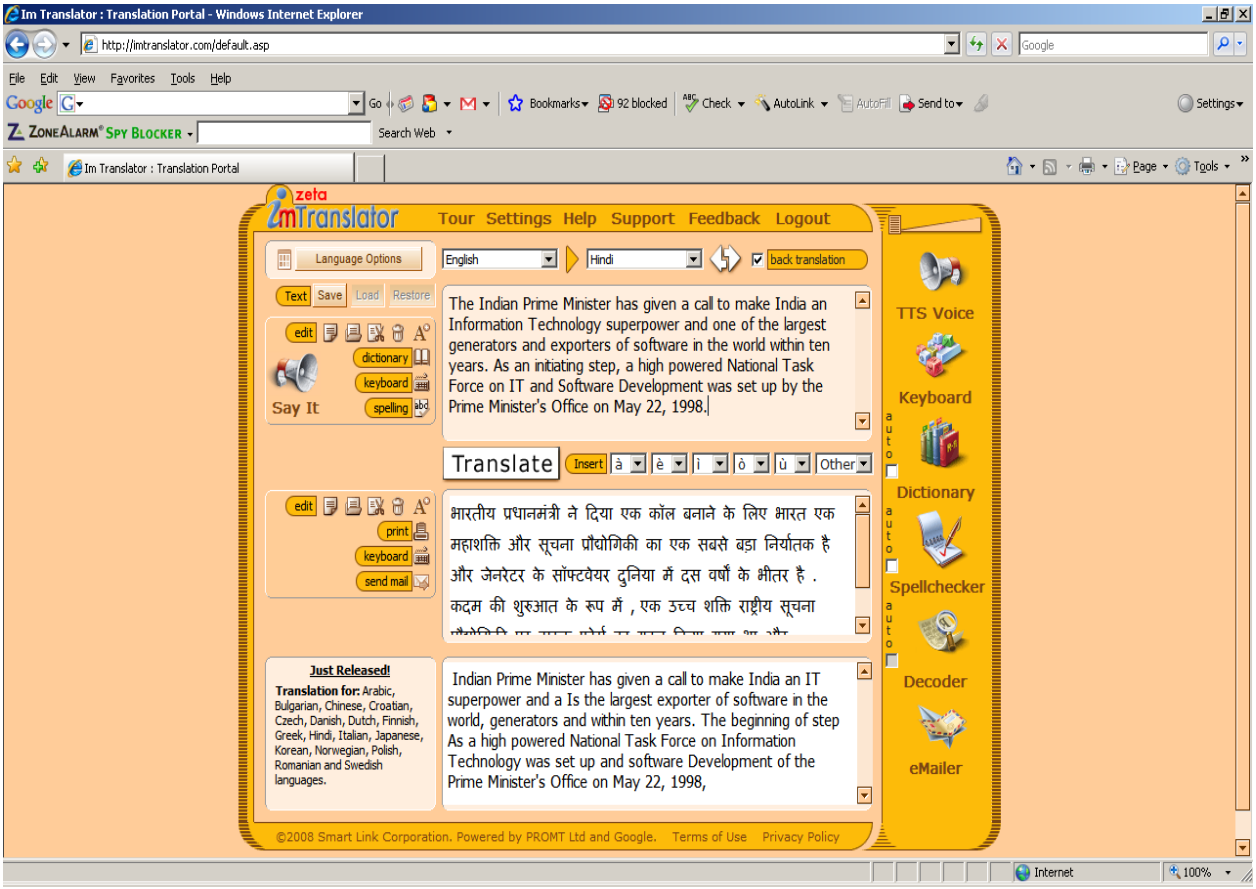
International Business Machines (IBM) has invested in MT technology to translate its documentation. Besides multinational corporations, some other major users of MT systems are the European Commission, the intelligence services, such as the Central Intelligence Agency (CIA), and the National Air Information Center in the USA.

There are several free MT software programs available on the internet. We focus on a few of these which convert text written in English to text in Hindi. Hindi is known to be one of the world's most widely spoken languages, specifically in the subcontinent of India. Approximately, there are about 500 million native Hindi speakers and the total number of people worldwide who understand the language could be as high as 800 million. Hindi is written in an easy to learn phonetic script called Devanagari which is also used to write Sanskrit, Marathi and Nepali.

ImTranslator (available at http://imtranslator.com/default.asp) is free online multilingual MT software which performs real-time English to Hindi voice and text translation as well as other

language translations such as English to Arabic, English to Chinese, etc. A user of the ImTranslator can enter multilingual text, check it for correctness, change unreadable messages, translate in different languages, look up words, and print results. The result of the translation using this translator is comprehensible and the meaning is conveyed though it is not 100% correct. There is also a "back translation" option available which allows the user to check if the meaning of the translated text is consistent with that of the original text. Also, the ImTranslator has a keyboard and users can type text using this keyboard in approximately twenty-five different languages (e.g., Hindi, Arabic, French and German). A user using this multilingual keyboard can type in text and using the translate option and get the text converted to the language of his/her choice. A screenshot along with the source (original) and the translated (target) text is shown in Figure 1.

Figure 1: Example of an imTranslator.



Another free online MT software program is Google's language translator, Google Translate (available at http://google.com/translate_t#) which is widely considered to be the best free online translator available. This translator can convert text as well as web page text contents written in about twenty-five different languages. Initially, Google's language translator was based on SYSTRAN. Recently, Google developed Google Translate using a statistical approach. In this approach, computer is provided with billions of words of text (input), both monolingual text in the target language, and aligned text consisting of examples of human translations between the

languages. Then, statistical learning techniques are applied to build a translation model. An example of the translation (English to Hindi) performed by Google Translate is shown in Figure 2.

Figure 2: Google Translate Page.

Yahoo's Babel Fish (available at http://babelfish.yahoo.com/) is yet another free online machine translator. Similar to Google's translator, Babel Fish can translate text as well as WebPages. But, Babel Fish does not have the capability to translate source text written in English to Hindi (target language).

OPERATIONAL MODELS

The use of advanced technologies (for example, IR, IE, MT) to extract, translated and present pertinent information to management will be discussed in this section. We show how these technologies can be combined to provide current information to enhance decision making. We propose possible combinations of these technologies. A simple process is shown in Figure 3.

Figure 3: A Generalize Process of Performing ES Using IE, and MT.

English Query: Asian stocks markets

Returned links:

Asian stocks tumble as US bailout optimism wanes
Channel News Asia, Singapore - 19 hours ago
Japan, **Asia**'**s** biggest **stock market**, and Australia both fell around 1.7 percent. Hong Kong and South Korea tumbled 1.5 percent, while Singapore also ended in **...** all 6,896 news articles » FNM - FRE

Asian stock market summary
Forbes, NY - 21 hours ago
The Nikkei 225 **Stock** Average ended 1.8 percent lower at 12400.65, as investors moved to take profits after the Nikkei gained 412 points in the previous **...**

Asian Stocks Drop for Second Day, Led by Materials Producers
Bloomberg - 30 minutes ago
10 (Bloomberg) -- **Asian stocks** fell for a second day, led by materials and shipping companies, after metals prices declined and cargo rates slumped on **...**

Select a document

HK shares open 1.8 pct lower led by resources - Tue Sep 9, 2008 9:55pm EDT
HONG KONG, Sept 10 (Reuters) - Hong Kong shares will open 1.8 percent lower on Wednesday with energy stocks walloped by retreating crude oil prices, which also dragged down shares in other commodity-linked stocks.
The benchmark Hang Seng Index .HSI is indicated to open down 376.25 points at 20,114.86.
The China Enterprises Index .HSCE of top locally listed mainland Chinese firms was set to open 1.8 percent lower.http://www.reuters.com/article/hongkongMktRpt/idUSHKF07948920080910

Information Extraction System

Financial Item: Hong Kong shares
Date/Time reported : Sep 10, 2008
Source: Reuters
Financial Quarter: Third-quarter
Financial Status: lower
Financial Fact: Fact
%Change: 1.8
…

Machine Translation System

Google Translator Translation: English » Hindi
वित्तीय आइटम : हांगकांग के शेयरों
तिथि / समय की रिपोर्ट : सिपाही 10 , 2008
स्रोत : रायटर
वित्तीय क्वार्टर : तीसरी तिमाही
वित्तीय स्थिति : कम
वित्तीय तथ्य : तथ्य
% बदलें : 1,8

ImTranslator:

वित्तीय आइटम : हांगकांग sharesDate / समय की रिपोर्ट : 10 सिपाही , 2008Source : ReutersFinancial क्वार्टर : तीसरी quarterFinancial स्थिति : lowerFinancial तथ्य : % तथ्य बदलें : 1,8

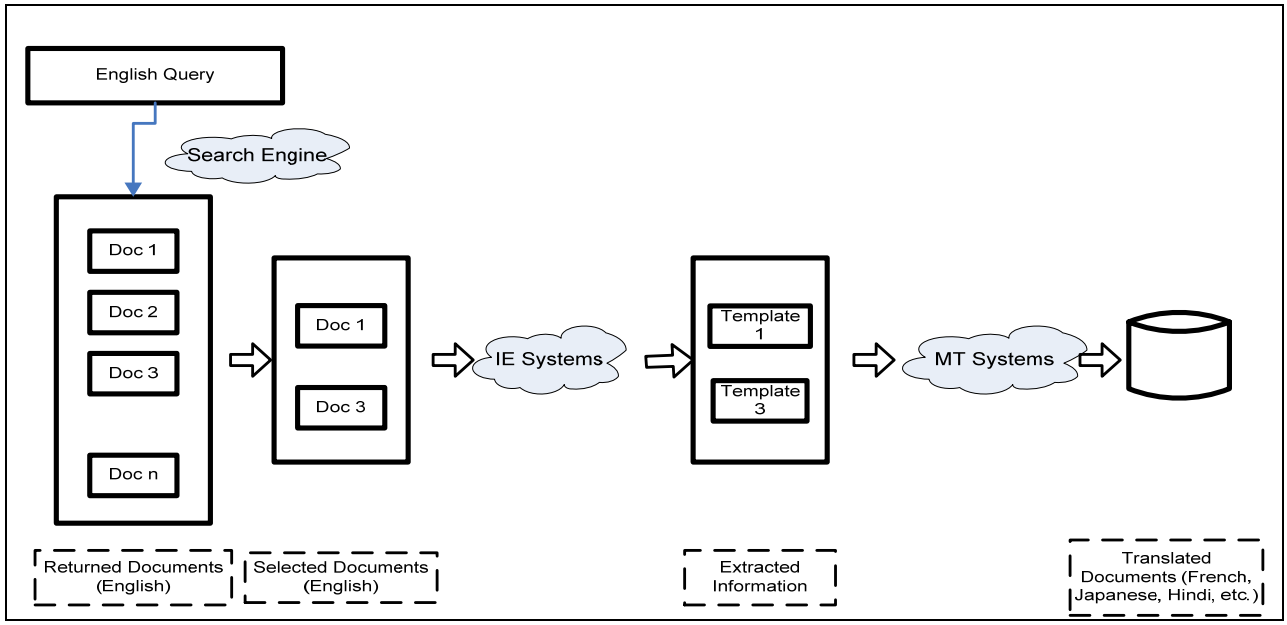
A query is submitted to the website of interest to search for the documents. The returned documents are input into an IE system to produce frames. The MT translates the frames into a target language. In the following paragraphs, the accuracy of the results of MT outputs obtained using Google's Google Translate and ImTranslator are discussed.

The translation (English to Hindi) of Google Translate was reviewed by a native Hindi speaker (one of the authors). Based on his assessment, the translation was accurate. The translated text conveyed the same meaning as the original text. Further, based on the native Hindi speaker's assessment, the translation done by ImTranslator was not very accurate. The translated text did not convey the same meaning as the original text since some words as well as phrases in the translated text had no relation to the original text or phrase. Also, some words in the translated text were not converted to Hindi.

In this proposed model (Figure 4), the query is input to the web browser in the source language (e.g., English). A list of items in English is returned. The user selects the documents of interest and sends each of them to the IE system for information extraction. The English frames are generated. These frames are then translated by the MT system into the target language (e.g., Hindi, French, Japanese, and Farsi) which can be stored in the database.

Figure 4: The English Query with IE and MT.

This model is useful for multinational corporations, whose managers might not be native English speakers, but have some knowledge of the concepts or expressions in English. Users input the



queries in English, to a web browser. The returned documents (in English) are passed to the IE system which generates output as frames for each document of interest. These frames are translated by an MT system to the target language (or language of choice) which can later be inserted into the database.

Extracting ES information using Text mining techniques

In this section, we discuss how to extract the information required by multinational corporations that was identified in Muralidharan 2003 (see the motivation section of this paper for details). The documents we use were specific to a country (e.g., India) and related to one of the three

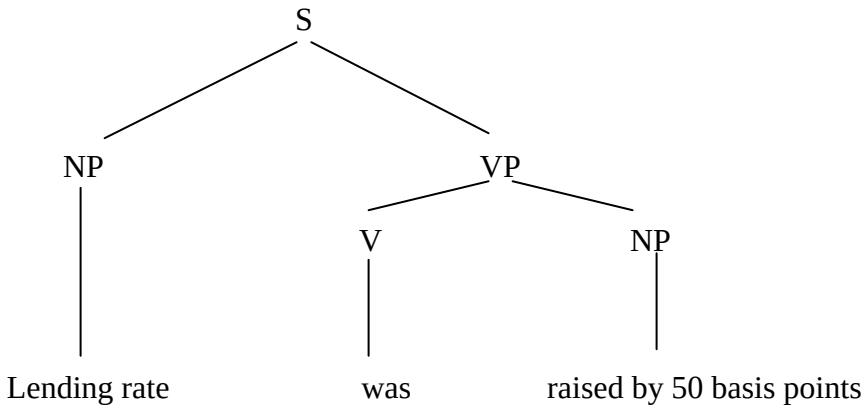
areas (for example, financial ES) identified by Muralidharan. Further, we used the Reuters website as our main source to retrieve information.

The FIRST (**Flexible Information extRaction SysTem),** was built using the knowledge engineering approach and its uses are diverse. FIRST's primary function was to extract financial information from the online articles. However, we have improved the system to extract other types of information, specifically those that match the areas mentioned by Muralidharan. Please refer to Conlon, Lukose, Hale, and Viknjamur (2007) for details of the implementation of the system.

In this discussion, we concentrate on how to extract Prime Interest Rates at various banks in India. The Central Bank sets the rate it lends to banks, following which individual banks adjust their rates to lend to businesses and consumers. We searched the Reuters website using "Indian Prime lending rates". We also searched related links on the pages that were returned after the initial search. The words appearing in these documents were analyzed using linguistic analysis techniques and information was extracted from them.

In general, a sentence consists of a noun phrase and a verb phrase. A brief syntactic structure for the sentence, "The central bank raised key lending rate," for example, is shown in Figure 5:

Figure 5: A Sample brief Syntactic Structure of a Sentence.



We use Lingua::EN::Tagger to identify noun phrases so that we can select the ones we are interested in.

Several noun phrases that appeared in the documents we looked at gave us clues to search for interesting information. 'Prime rate' and rates were of particular interest. We also found that these documents contain many common verbs such as "increase," "hold," "raised," "decrease," "consider," and "lowered," etc.

The word patterns of several sentences in the extracted documents were of interest and showed similarities. To test and demonstrate these patterns, we extracted each word from these documents and inserted them into a database. Subsequently, each word was inserted into a column such that fifty columns were generated per row/tuple. Each row represented a sentence.

SQL statements were used to identify the patterns we found interesting. Some sample records that contains the key terms we are interested, "rate," are shown in Figure 6.

Figure 6: Sample Records that Contain the Term "rate".

Subject Noun	Verb	Objective Noun Phrase
rate	could	change and you……..
rate	gained	to 6.62 percent……
rates	averaged	6.52 percent.
rates	climb	to 6.62 percent;
rates	on	15-year fixed-rate loans averaged……
rates	rise	for first time in four weeks……..
rates	rose	for the first time in…….
rates	ticked up	a little in the……

The noun phrases "rate/s" appear in front of the key verbs. The main interest is the subject of the sentence and the noun phrases that appear after them. Thus, these noun and verb phrases as well as the object noun phrases helped us identify relevant information from the documents.

In addition to noun phrases and verb phrases, the adjectives also provided important information. For example, we found some adjectives that express a bank's decision on lending rates. Adjectives that appeared in noun phrases like "raise rate(s)," and "lower rates(s)," etc., allowed us to find information about changing of lending rates of banks. Some examples of the adjectives from Reuters articles are shown below.

- India's biggest mortgage firm Housing Development Finance Corp Ltd (HDFC.BO: Quote, Profile, Research, Stock Buzz) and leading private bank ICICI Bank (ICBK.BO: Quote, Profile, Research, Stock Buzz) said on Monday they were ***raising*** lending and deposit rates.

- HDFC said its prime lending rate would go up by 50 basis points from Tuesday, and ICICI said rates on consumer loans would ***rise*** by 75 basis points on Tuesday.

- The increases come a week after the central bank ***raised*** its key lending rate by 50 basis points to 8.5 percent, its highest since March 2002 and the second increase this month, to rein in inflation, which has surged to 13-year highs.

As seen above, some terms express the quantity of the subjects and the objects of the sentences. These include "biggest," "second," and the more precise expression would be a number followed by the word "percent" or the percent sign, e.g., 50% or a number followed by "basis points")e.g. 50 basis points). Two sample sentences are shown below.

- The increases come a week after the central bank raised its key lending rate by ***50 basis points*** to ***8.5 percent***,

- Corporation Bank (CRBK.BO: Quote, Profile, Research, Stock Buzz) and Union Bank of India (UNBK.BO: Quote, Profile, Research, Stock Buzz), have also raised rates by *50 basis points* in recent days.

Some adjectives such as "large" and "small" indicate the degrees of precision of the change of the verb. Expressions of similar meaning were used frequently and additional terms added to the key words of interest. For example, the word "lowered" has similar meaning to the phrases "went down," "ticked down," and "decreased." Also, the noun "rates" can be called "prime lending rate," "loan APR." Some linguistic information about word synonyms, antonyms, hyponyms, etc. can be found from some important sources such as WorldNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990).

From the extracted data set we were able to find patterns for rule extractions. For example, if we are concerned about the change of interest rates in foreign countries, particularly India, we looked for the phrase "lending rate" and the terms that appear around them. Some portions of the examples we found from the articles we used are:

key	lending rate	is 8.5 percent
India's HDFC Bank raised its prime	lending rate	by 75 basis points to 16 percent
the central bank raised its key	lending rate	by 50 basis points to 8.5 percent
State Bank of India, raised its prime	lending rate	by 50 basis points to 12.75 percent last Thursday. 6

We were able to develop a set of rules for information extraction. These rules can be used to extract information from online documents. To illustrate this, we show a portion of a document from the Reuters news website http://www.reuters.com/article/rbssBanks/idUSDEL19870520080630.

NEW DELHI, June 30 (Reuters) - State Bank of India (SBI.BO: Quote, Profile, Research, Stock Buzz), the country's largest lender, may have to set aside 10 billion rupees ($230 million) to cover depreciation of its treasury portfolio this fiscal year, its chairman said on Monday. Indian banks have seen the value of their treasury holdings eroded by rising bond yields, which have risen to their highest in seven years as inflation has surged to 13-year highs and the central bank has tightened policy to rein it in.
"It will be a significant amount, but we will know it only at the end of March," Chairman O.P. Bhatt told reporters when asked how much SBI would set aside. "It can easily be 10 billion rupees."
Bhatt said he hoped to maintain SBI's net interest margin at 3 percent in the fiscal year ending March 2009, the same as the past two years.
The central bank last week raised its key lending rate by 50 basis points to 8.50 percent, the second rise in June taking the rate to a six-year high. It also increased banks' reserve requirements by 50 basis points
…

Some extracted information can be found from the above document is:

Date: June 30
City: New Delhi
(object phrase): The central bank
Details: raised key lending rate
Amount (subject phrase): by 50 basis points

Finally, a MT system can translate these items into a language of interest (Hindi, in this research) such as:

Translation: English » Hindi

तिथि : 30 जून
शहर : नई दिल्ली
(वस्तु वाक्यांश) : केंद्रीय बैंक
विवरण : उठाया मुख्य उधार दर
राशि (वाक्यांश का विषय है) : द्वारा 50 बिंदुओं के आधार पर

We have demonstrated that the extracted information makes it possible to search for "changes in lending rates". Further, a large amount of this type of information helps to formulate the general information for ES.

CONCLUSION

Collecting information from the business environment is made easier with the use of ES. The increasing amount of business information that is available in textual format and in many languages makes the use of ES imperative to enable timely decisions by management. Improving the decision-making process for managers was the purpose of this research. Using our technique/tool, managers can easily understand local business information since such information can be converted to the language of interest. We used linguistic analysis to aid in our scanning and extraction of information. The international business decision-makers, governments, etc. are the beneficiary of the output of our tool and they will receive comprehensive information more quickly and hence make better and faster decisions. Thus they can create a competitive advantage with the use of our tool.

REFERENCES

Aguilar, F. J. (1967). Scanning the Business Environment, Macmillan *Co*., New York.

ALTAVISTA bablefish - Retrieved April 16, 2007 from http://world.altavista.com/

ANNIE (A Nearly-New IE system). Retrieved March 16, 2007, from http://gate.ac.uk/ie/annie.html

Appelt, D., & Israel, D. (1999, August). Introduction to information extraction technology. Retrieved February 19, 2006, from http://www.ai.sri.com/~appelt/ie-tutorial/

Cardie, C. (1997). Empirical methods in information extraction, *AI Magazine, 18*(4), 65-80.

Chen, H., Chau, M. & Zeng, D. (2002). CI-Spider: a Tool for Competitive Intelligence on the Web, *Decision Support Systems*, 34(3), 1-17.

Conlon, S., Lukose S., Hale J., & Viknjamur A. (2007). Automatically Extracting and Tagging Business Information for E-Business Systems Using Syntactic and Semantic Analysis, *Semantic Web Technologies and eBusiness: Virtual Organization and Business Process Automation*. A. F. Salam & Jason Steven edited. Idea Group Inc, 101-126.

Cowie, J., & Lehnert, W. (1996). Information extraction. *Communications of the ACM, 39*(1), 80-91.

Culnan, M. J. (1983). Environmental Scanning: The Effects of Task Complexity and Source Accessibility on Information Gathering Behavior, *Decision Sciences***,** 14(2), 194-206.

Daft, R. L., Sormunen, J., & Parks, D., 1988, Chief Executive Scanning, Environmental Characteristics, And Company Performance: An Empirical Study, *Strategic Management Journal*, 9(2), 123-139.

Decker, R., Wagner, R. & Scholz, S. W. (2005). An Internet-based Approach to Environmental Scanning and Marketing Planning, *Marketing Intelligence and Planning*, 23(2), 189-199.

Dollinger, Marc J. (1984). Environmental boundary spanning and information processing effects on organizational performance, *Academy of Management Journal*, 27(2), 351-368.

Fuld & Company. (2007). Technology Risk and Reward - Intelligence Software Report® 2006/2007, Retrived on April 6, 2007 from http://www.fuld.com/Products/ISR2006/HomePage.html

GATE (A General Architecture for Text Engineering). Retrieved March 16, 2007, from http://gate.ac.uk/

Google Translator- Retrieved August 08, 2008 from http://www.google.com/language_tools

Hambrick, D. C. (1982). Environmental Scanning and Organizational Strategy, *Strategic Management Journal*, 3(2), 159-174.

He, S. & Zhu, J. (2007). Corporate Blogs of 40 Fortune 500 Companies: Distribution, Categorization and Characteristics, *Journal of International Technology and Information Management*, 16(3), 45-58.

Jacobs, P. S. & Rau, L. (1990). SCISOR: Extracting information from on-line news. *Communications of the ACM, 33*(11), 88-97.

Kalczynski, P. J. (2005). Time Dimension for Business News in the Knowledge Warehouse, *Journal of International Technology and Information Management*, 14(3), 21-32.

Keegan, W. (1974). Multinational Scanning: A study of the Information Sources Utilized by Headquarters Executives in Multinational Companies, *Administrative Science Quarterly*, 19, 411-421.

Lingua::EN::Tagger, retrieved August 15, 2008, http://search.cpan.org/~acoburn/Lingua-EN-Tagger-0.15/Tagger.pm.

Liu, S., Turban, S. & Lee, M. K. O. (2000). Software Agents for Environmental Scanning in Electronic Commerce, *Information Systems Frontiers*, 2(1), 85-98.

Manning C., & Schutze H. (2002). *Foundations of statistical natural language processing*, (5th ed.). Cambridge, MA: The MIT Press.,

McManus, D. J. & Snyder, C. A. (2003). Knowledge Management: The Role of EPSS, *Journal of International Technology and Information Management,* 12(2), 17-28.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography, 3*(4), 235-244.

Muralidharan, R. (2003). Environmental Scanning and Strategic Decisions in Multinational Corporations, *Multinational Business Review*, 11(1), 67-88.

Newgren, K. E., Rasher, A. A. & LaRoe, M. E. (1984). An Empirical Investigation of the Relationship Between Environmental Assessment and Corporate Performance. In Proceedings of *Proceedings of the 44th Annual Meeting of the Academy of Management,* held in Washington, DC, edited by John A. Pearce II and Richard B. Robinson Jr, 352-356.

Ngamkroeckjoti, C. & Johri, L. M. (2003). Coping with Hypercompetition in the Financial Services Industry in Thailand: Environmental Scanning Practices of Leaders and Followers, *The International Journal of Bank Marketing,* 20(6), 359-368.

Pirolli, P. & Card, S. (1999). Information Foraging, *Psychological Review*, 106(4), 643-675.

Salton, G. (1970). Automatic processing of foreign language documents, *Journal of the American Society for Information Science*, 21(3), 187–194.

Subramanian, R., Fernandes, N. & Harper, E. (1993). Environmental Scanning in US Companies: Their Nature and Their Relationship to Performance, *Management International Review*, 33(3), 271-285.

Reuters News – Retrieved July 24, 2008 from http://search.us.reuters.com

SYSTRAN - Retrieved April 16, 2007 from http://www.systransoft.com/

West, Joseph John. (1988). *Strategy, Environmental Scanning, and Their Effect Upon Firm Performance: an Exploratory Study Of the Food Service Industry* . Ph.d. Thesis, Virginia Polytechnic Institute and State University.

This Page is Blank