

The Trend of Knowledge Management in e-Business Research (2000-2009): A Text Mining Analysis

Shianghau Wu and Jiannjong Guo

Abstract—In recent years, much research has been devoted to the knowledge management in e-business; however, few have deployed the study of the research trend. In fact, grasping the research trend is imperative for enterprises and researchers. The study aims at analyzing the trend change of the knowledge management in e-business from 2000 to 2009. The contribution of the study includes the following two points. First, the study has found the shift of the knowledge management in e-business research from the customers and community focus to the topic of organization. Second, the study finds a new way of literature survey by using the text mining method in order to explore the trend change.

Index Terms—Knowledge management; text mining; trend; konstanz information miner(KNIME)

I. INTRODUCTION

Knowledge Management has become an important research focus. Many e-business enterprises began to apply different knowledge *management* methods. So grasping the trend of the knowledge management application to e-business research is imperative to researchers and enterprises.

In this study, the author applies to the text mining analysis to achieve this goal. The study analyzes the abstracts of academic papers of the knowledge management in e-business on major electronic databases from 2000 to 2009 in order to get keywords and grasp the trend during the time frame.

The rest of the paper is organized as follows. First, the study begins with the literature review of the knowledge management in e-business. Second, the overall research design is outlined, the research sites are described and data collection methods are exposed. Third, the main results are then presented and discussed. Fourth, the study uses the comparative method to analyze the difference of the text mining results and then induces the trend of the knowledge management in e-business researches. The paper concludes with implications and future research avenues.

II. LITERATURE REVIEW

For enterprises, adopting e-business is imperative to the development. Knowledge management belongs to one of the internal factors involved in the first stage of decision making. In the second stage of achieving e-business success, factors

related to the ongoing knowledge management also affects the successful implementation of the e-business[1]. As for the knowledge management application and e-business adaptation in small and medium enterprises, scholars have mentioned cost factors, organizational and cost factors as the driving force for achieving success[2][3]. Other views suggest the strategic factor is also the driving force for small enterprises to apply to knowledge management systems in e-business[4], and du Plessis and Boon (2004) also refer to the customer relationship management as another important factor to facilitate knowledge management application to e-business[5].

III. METHODOLOGY

The study purposes to use the text mining method to analyze the trend of the green supply chain management. Text mining is one of the data mining methods, which learn from samples of past experience. In the text mining method, the text will be processes and transformed into a numerical representation. Fig.1 represents the document classification application and shows the process of categorization from originally unstructured texts [6]. It is the ordinary process of the text mining method.

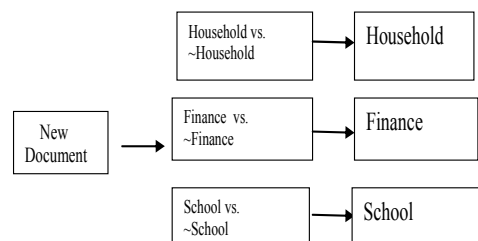


Fig.1. Text categorization (weiss et al, 2005)

The text mining method is widely applied to information management on websites, biological data and customer relationship management. Other applications mainly relate to library management and literature critique. For example, Clement (2008) uses the text mining method to analyze Gertrude Stein’s “The Making of Americans” and has different conclusions from that of other literature critics [7]. And Trappey (2007) applies the text mining method to analyze patent brochures. As for the studies of green supply chain management, many analysts were made up to the present. But few analyses were related to text mining [8].

A. Research Design

The text mining method of the study is implemented by means of the text processing function of the KNIME software. The KNIME (Konstanz Information Miner) software is the open source software created by the Konstanz University in Germany. The study uses the KNIME software to analyze the knowledge management application to e-business related papers on the ProQuest and

Manuscript received April 15, 2012; revised May 20, 2012.
 Shianghau Wu is with Faculty of Management and Administration
 Macau University of Science and Technology,
 Taiwan(shwu@must.edu.mo).
 Jiannjong Guo is with Tamkang University, Taiwan.

Electronic Journal Service (EBSCO) from 2000 to 2009 and then uses the comparative method to explore the changes of keywords of academic paper abstracts between 2000 and 2009. The standard of choosing academic papers is based on the relevance to the knowledge management in e-business.

The KNIME text processing plugin follows the procedures as follows [9]:

- Parsing: The text and the structure were extracted and represented in a data structure.
- BoW(Bag of Words) Creator: This node creates a bag of words (BoW) of a set of documents.
- Tags to String: this node converts the term's tag values of the specified tag types to strings. For each selected tag type a column is appended, containing the strings representation of the corresponding tag value.
- Kuhlem Stemmer: This node allows the user to reduce terms to their stem. The study used stemming algorithm is the Kuhlen stemmer.
- N Chars Filter: The node filters terms consisting of less than a specified number. In the study, the specified number is set as four in order to filter out definite and indefinite articles (e.g., a, an, the).
- Number Filter: The node filters all terms consisting of numbers only.
- The Porter Stemmer: The node reduces the words to their stems according to Porter (1997)[10].
- The Punctuation Erasure: The node removes all punctuation marks.
- Stop Words Filter: The node filters all stop words.
- Keywords extraction: After the above-mentioned processes, all irrelevant terms can be filtered out. To compute these frequencies two nodes are available, one to compute the well known term frequency tf, relative or absolute, and one to compute the inverse document frequency idf. These nodes require a bag of words as input data table, and return a bag of words as output data table, with additional columns containing the computed frequencies. It is possible to specify minimum and maximum values or a particular number k, so that only the k terms with the highest frequency are kept. The study uses the default k value on the KNIME Text Processing Plugin. The study also utilizes the “Chi-square keyword extractor” [11] and “Keygraph keyword extractor” [12] to find keywords.
- Transformation: The study applies the “Term Vector” node to create a term vector for each term in order to make further application, such as the neighborgram.
- Visualization: The study applies the “Tagcloud” node. It provides a typical tagcloud with some additional options, such as different arrangements of terms, i.e. size sorted, alphabetic, inside out, adjustable minimal and maximal font size, or transparency of terms etc. The node requires a input data table consisting of a bag of word with a additional column containing a weight or score of each term.

According to these processes, the research design in this study can be shown in Fig. 2, which is presented in the KNIME software. Each process is shown in nodes on the graph.

It is worthwhile to mention the study also applies the neighborgram node originated from Berthold et al.

(2005)[13]. The aim of the neighborgram is to allow users to interact with the clustering process by selecting, discarding, or fine-tuning potential cluster candidates. The advantage of the neighborgram (neighborhood histogram) is its interpretation.

The neighborgram in this study is made by fuzzy clustering by the Rule Engine process as shown on Fig. 2 (Node 18).

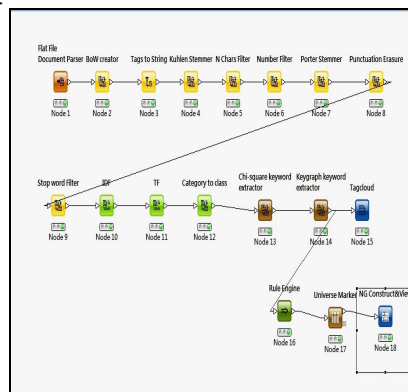


Fig. 2. Research design in the study

B. Data Collection

The study asorts the data from the green supply chain management related papers’ abstracts on two major academic searching engines (ProQuest and EBSCO). The time frame is from 2000 to 2009. The goal of this study is to compare the keywords of these academic papers’ abstracts from 2000 to 2009 in order to grasp the trend of research in these ten years. The study analyzes 130 academic papers’ abstracts from 2000 to 2009.

C. Descriptive Analysis

Descriptive analysis offers first insights into the research material. The distribution of the publication is shown in Fig. 3.

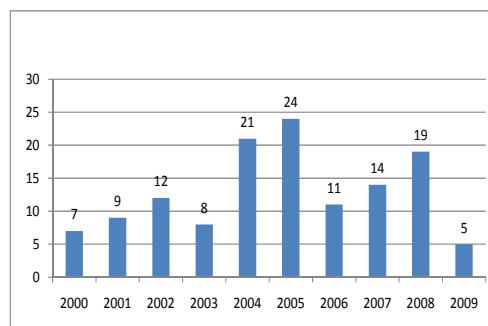


Fig. 3. The distribution of the papers across the analyzed period (2000-2010)

The analyzed papers are all in English. It is due to the fact that the KNIME software can only analyze English texts. In Fig. 3, particularly large volume of papers is found in 2004 and 2005. The reason is that academic papers in the EBSCO searching engine are of larger volume in 2004 and 2005, so the numbers of papers are expected to be considerably higher.

IV. RESULTS AND DISCUSSION

A. Tag Clouds

The text mining method in the study represents a valuable

contribution. The tag clouds results are shown in Fig. 4.

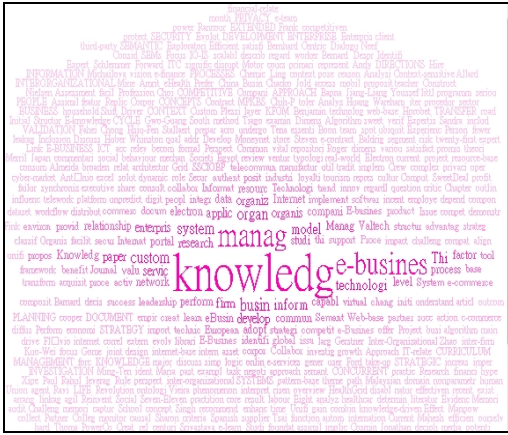


Fig. 4. Tag cloud of keywords from 2000 to 2009

On Fig. 4, we would know the important keywords of academic paper abstracts from 2000 to 2009 other than “knowledge”, “management” and “e-business” are strategies, adopt, leadership, customers, community and organization.

B. Neighborgrams

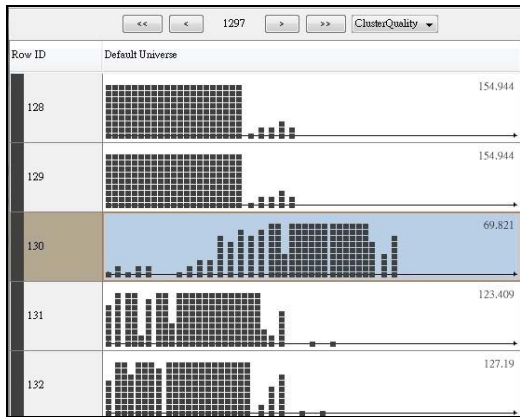


Fig. 5. Neighborgram of keywords from 2000 to 2009

The study also checks the relationship of each keyword by analysing two neighborgrams from 2000 to 2009 text data. Neighborgrams are made for any object of the dataset or in the case of very large volume datasets. A single neighborgram reflects the data distribution around a certain object, the centroid of the neighborgrams [14]. The neighborgram is shown in Fig. 5.

On Fig. 5, numbers in each cell represent the distance to the centroid [9]. These two graphs show the data distribution of each row (keyword). Due to the larger volume of data, Fig. 5 only presents the partial portion of neighborgram. Row 130 on Fig. 5 represents the keyword ‘knowledge’. We can observe the data distribution of ‘knowledge’ is much different from other rows and with closer vicinity to the centroid. The distance to centroid of the keyword ‘knowledge’ is 69.821.

The study attempts to find the closer keywords to ‘knowledge’ and get the results in the following tables. It is because we cannot get the actual distance among keywords from Fig. 4. The criteria of searching closer keywords are the distance to the centroid.

The study concludes the results from tag clouds (Figure 4) and neighborgrams (Fig. 5) and finds the keywords which are closer to the keyword ‘knowledge’ in Table I.

TABLE I: CLOSER KEYWORDS TO THE KEYWORD ‘KNOWLEDGE’

Keywords	Distance to the Centroid
adopt	65.633
organization	73.925
community	79.458
customers	99.167

C. Keywords Life Cycle

In order to find the trend of five keywords in Table I concluded from the tag cloud and the neighbourgram, the study attempts to depict the keyword life cycle graph. The standard of determining the trend of each keyword in each year is the score calculated from “Keygraph Keyword Extractor” (Node 14 in Figure 2). The partial result graph of the “Keygraph Keyword Extractor” is shown in Fig. 6.

Row ID	Keyword	Score	Document
90	Valtech	19	"Organis logic busin exampl knowled g mana
91	logic	14	"Organis logic busin exampl knowled g mana
92	technolog	11	"Organis logic busin exampl knowled g mana
93	busin	9	"Organis logic busin exampl knowled g mana
94	firm	7	"Organis logic busin exampl knowled g mana
95	consult	6	"Organis logic busin exampl knowled g mana
96	Organis	6	"Organis logic busin exampl knowled g mana
97	Despr	6	"Organis logic busin exampl knowled g mana
98	manag	6	"Organis logic busin exampl knowled g mana
99	exampl	6	"Organis logic busin exampl knowled g mana
100	knowledg	26	"Eight question custom knowled g manag e-b
101	custom	24	"Eight question custom knowled g manag e-b
102	e-business	8	"Eight question custom knowled g manag e-b
103	manag	8	"Eight question custom knowled g manag e-b
104	cultur	6	"Eight question custom knowled g manag e-b
105	commu	6	"Eight question custom knowled g manag e-b

Fig. 6. The partial result graph of “keygraph keyword extractor”

According to Fig. 6, each keyword has its corresponding score to represent the importance.

Five keywords life cycle graphs are shown from Fig. 7 to Fig. 10. The vertical axis in each graph represents the score of each keyword.

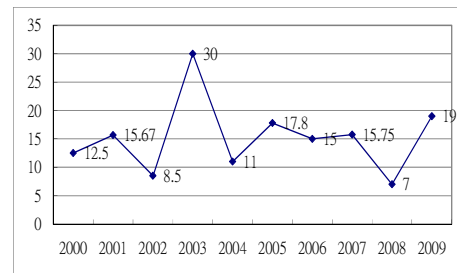


Fig. 7. The keyword life cycle of the keyword “organization”

According to Figure 7, the study finds the keyword “organization” scores highest in 2003. But the trend is stable in other years, except the decreasing trend in 2008.

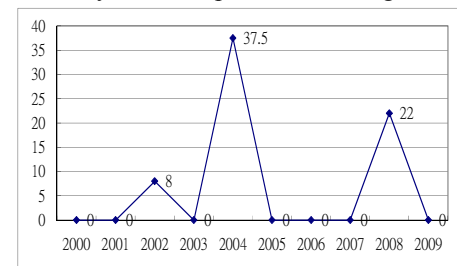


Fig. 8. The keyword life cycle of the keyword “community”

According to Fig. 8, the study finds the keyword “model” scores only in 2002, 2004 and 2008. The keyword “community” does not have a clear live cycle according to Fig. 8.

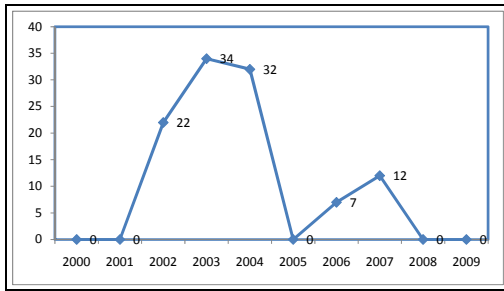


Fig. 9. The keyword life cycle of the keyword “customers”

According to Fig. 9, the keyword “customers” is in the increasing trend in 2002, 2003, but in the decreasing trend from 2003 to 2005. And the score bounces from 7 to 12 from 2006 to 2007.

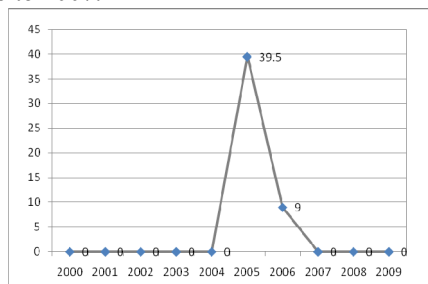


Fig. 10. The keyword life cycle of the keyword “adopt”

According to Fig. 10, the scores of the keyword “adopt” are zero from 2000 to 2004, except the years of 2005 and 2006. The keyword “adopts” scores highest in 2005, but then decreases.

V. RESEARCH LIMITATION

The research limitation of the study is the less volume of analyzed papers. Although the standard of choosing academic papers is based on the relevance of the knowledge management in e-business, the analyzed papers can't include all published English academic papers. However, the study concludes the trend of five major keywords from 2000 to 2009. It offers some useful information for further research.

VI. CONCLUSION

The contribution of the study is the development of new literature survey method to identify the trend of the knowledge management application in e-business research from 2000 to 2009. From the study, the research focuses of knowledge management in e-business shifts from customers and community to organization. The study offers more

insights for related research.

REFERENCES

- [1] J. Xu, and M. Quaddus, “Achieving Sustainable E-Business Success: Development and Application of a Model of E-Business Adoption, Success and Sustainable Success,” in *E-Business in the 21st Century: Realities, Challenges and Outlook*, World Scientific, pp. 415-416, 2009.
- [2] F. Bee and R. Bee, “Training Needs Analysis and Evaluation,” Institute of Personnel Management, London, 2004.
- [3] L. Jessup and J. Valacich, *Information Systems Today*, Prentice-Hall, Englewood Cliffs, NJ, 2003.
- [4] M. Fletcher and S. Harris, “Seven aspects of strategy formation: exploring the value of planning,” *International Small Business Journal*, vol. 20 no. 3, pp. 297-312, 2002.
- [5] M. D. Plessis and J. A. Boon, “Knowledge management in eBusiness and customer relationship management: South African case study findings,” *International Journal of Information Management*, vol. 24, no. 1, pp. 73-86, 2004.
- [6] S. Weiss, N. Indukhya, T. Zhang, and F. Damerou, *Text Mining: Predictive Method for Analyzing Unstructured Information*, Springer, 2005.
- [7] T. E. Clement, “A thing not beginning and not ending: using digital tools to distant-read Gertrude Stein’s *The Making of Americans*,” *Literary and Linguistic Computing*, vol. 23, no. 3, pp.361-380, 2008.
- [8] A. J. C. Trappey and C. V. Trappey, “An R &D knowledge management method for patent document summarization,” *Industrial Management & Data Systems* vol. 108, no. 2, pp. 245-257, 2007.
- [9] K. Thiel. (2009). The KNIME Text Processing Plugin. [Online]. Available: <http://www.knime.org>.
- [10] M. F. Porter, *An Algorithm for Suffix Stripping*, Computer Laboratory, Cambridge, pp.313-316, 1997.
- [11] Y. Matsuo and M. Ishizuka, “Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information,” *International Journal on Artificial Intelligence Tools*, vol. 13, no.1, pp. 157-169, 2004.
- [12] K. Thiel, “KeyGraph: automatic indexing by co-occurrence graph based on building construction metaphor,” *IEEE International Forum on Research and Technology Advances in Digital Libraries*, ADL 98, Proceedings, 1998.
- [13] M. R. Berthold, B. Wiswedel, and D. E. Patterson, “Interactive exploration of fuzzy clusters using neighborgrams,” *Fuzzy Sets and Systems*, vol. 149, no. 1, pp.21-37, 2005.
- [14] B. Wiswedel, *Lernen in Parallelen Universen, Dissertation zur Erlangung des akademischen Grades des Doktors der Naturwissenschaften*, Universitaet Konstanz, 2008.



Shianghau Wu is currently an Assistant Professor in the Faculty of Management and Administration at Macau University of Science and Technology. He received his Bachelors and Masters at the National Taiwan University in 1997 and 2001, respectively and his Doctorate at the National Chengchi University Taiwan, China in 2006. His current research areas cover economics, public administration and data mining applications.



Jiannjong Guo is currently the associate professor in Tamkang University in Taiwan. He received his Doctorate at the Sussex University in U.K. His current researches cover political economy and international investment.