Expert Systems with Applications 40 (2013) 3160–3173

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Knowledge management vs. data mining: Research trend, forecast and citation approach

Hsu-Hao Tsai *

Department of Commerce Technology and Management, Chihlee Institute of Technology, No. 313, Sec. 1, Wunhua Rd., Banciao District, New Taipei City 22050, Taiwan, ROC

ARTICLE INFO

Keywords:
Knowledge management
Data mining
Research trend analysis
Bibliometric methodology

ABSTRACT

Knowledge management (KM) and data mining (DM) have become more important today, however, there are few comprehensive researches and categorization schemes to discuss the characteristics for both of them. Using a bibliometric approach, this paper analyzes KM and DM research trends, forecasts and citations from 1989 to 2009 by locating headings "knowledge management" and "data mining" in topics in the SSCI database. The bibliometric analytical technique was used to examine these two topics in SSCI journals from 1989 to 2009, we found 1393 articles with KM and 1181 articles with DM. This paper implemented and classified KM and DM articles using the following eight categories—publication year, citation, country/territory, document type, institute name, language, source title and subject area—for different distribution status in order to explore the differences and how KM and DM technologies have developed in this period and to analyze KM and DM technology tendencies under the above result. Also, the paper performs the K–S test to check whether the distribution of author article production follows Lotka's law. The research findings can be extended to investigate author productivity by analyzing variables such as chronological and academic age, number and frequency of previous publications, access to research grants, job status, etc. In such a way characteristics of high, medium and low publishing activity of authors can be identified. Besides, these findings will also help to judge scientific research trends and understand the scale of development of research in KM and DM through comparing the increases of the article author. Based on the above information, governments and enterprises may infer collective tendencies and demands for scientific researcher in KM and DM to formulate appropriate training strategies and policies in the future. This analysis provides a roadmap for future research, abstracts technology trend information and facilitates knowledge accumulations, therefore the future research can concentrated in core categories. This implies that the phenomenon "success breeds success" is more common in higher quality publications.

1. Introduction

1.1. Knowledge management

Knowledge management (KM) does not carry its name accidentally because management normally means that 'something' has to be managed (Wiig, Hoog, & Spex, 1997). Since Polanyi's discussion of the distinction between explicit and tacit knowledge (Polanyi, 1966), researchers were developed a set of management definitions, concepts, activities, stages, circulations, and procedures, all directed towards dealing with objects in order to describe the framework of KM as the KM methodology. Different KM working definitions, paradigms, frameworks, concepts, objects, propositions, perspectives, measurements, impacts, have been described for investigating the question of: What is KM? What are its methods and techniques? What is its value? And what are its functions for supporting individual and organizations in managing their knowledge (Drew, 1999; Heijst, Spek, & Kruizinga, 1997; Hendriks & Vriens, 1999; Johannessen, Olsen, & Olaisen, 1999; Liao, 2002; Liebowitz, 2001; Liebowitz & Wright, 1999; Nonaka, Umemoto, & Senoo, 1996; Rubenstein-Montano et al., 2001; Wiig, 1997; Wiig et al., 1997; Wilkins, Wegen, & Hoog, 1997).

For example, the concept of 'the knowledge-creating company' is a management paradigm for the emerging 'knowledge society', and information technology can help implement this concept (Nonaka et al., 1996). Some articles have investigated issues concerning the definition and measurement of knowledge assets and intellectual capital (Liebowitz & Wright, 1999; Wilkins et al., 1997). A conceptual framework presents KM as consisting of a repertoire of methods, techniques, and tools with four activities performed sequentially (Wiig et al., 1997). These are also combined with another extension of KM working definitions and its historical development (Wiig, 1997). From the organizational perspective, corporate memories can act as a tool for KM on three types of learning in organizations: individual learning, learning through

* Tel.: +886 2 22576167; fax: +886 2 22521777.
E-mail address: simon705@mail.chihlee.edu.tw

0957-4174/$ - see front matter
http://dx.doi.org/10.1016/j.eswa.2012.12.029

direct communication, and learning using a knowledge repository (Heijst et al., 1997). Another example is innovation theory based on organizational vision and KM, which facilitates development-integration and application of knowledge (Johannessen et al., 1999). For strategy, Drew explores how managers might build KM into the strategy process of their firms with a knowledge perspective and established strategy tools (Drew, 1999). Furthermore, a systems thinking framework for KM has been developed, providing suggestions for what a general KM framework should include (Rubenstein-Montano et al., 2001). Also, the emergence and future of KM, and its link to artificial intelligence been discussed (Liebowitz, 2001). Knowledge inertia (KI), means stemming from the use of routine problem solving procedures, stagnant knowledge sources, and following past experience or knowledge. It may enable or inhibit an organization's or an individual's ability on problem solving (Liao, 2002). On the other hand, the organizational impact of KM and its limits on knowledge-based systems are discussed in order to address the issue of how knowledge engineering relates to a perspective of KM (Hendriks & Vriens, 1999). These methodologies offer technological frameworks with qualitative research methods and explore their content by broadening the research horizon with different perspectives on KM research issues.

1.2. Data mining

Data mining (DM) is an interdisciplinary field that combines artificial intelligence, database management, data visualization, machine learning, mathematic algorithms, and statistics. DM, also known as knowledge discovery in databases (KDD) (Chen, Han, & Yu, 1996; Fayyad, Piatetsky-Shapiro, & Smyth, 1996a), is a rapidly emerging field. This technology provides different methodologies for decision-making, problem solving, analysis, planning, diagnosis, detection, integration, prevention, learning, and innovation.

This technology is motivated by the need of new techniques to help analyze, understand or even visualize the huge amounts of stored data gathered from business and scientific applications. It is the process of discovering interesting knowledge, such as patterns, associations, changes, anomalies and significant structures from large amounts of data stored in databases, data warehouses, or other information repositories. It can be used to help companies to make better decisions to stay competitive in the marketplace. The major DM functions that are developed in commercial and research communities include summarization, association, classification, prediction and clustering. These functions can be implemented using a variety of technologies, such as database-oriented techniques, machine learning and statistical techniques (Fayyad, Piatetsky-Shapiro, & Smyth, 1996b).

DM was defined by Turban, Aronson, Liang, and Sharda (2007, p. 305) as a process that uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from large databases. In an effort to develop new insights into practice-performance relationships, DM was used to investigate improvement programs, strategic priorities, environmental factors, manufacturing performance dimensions and their interactions (Hajirezaie, Mohammad, Husseini, Barfourosh, & Karimi, 2010). Berson, Smith, and Thearling (2000), Lejeune (2001), Ahmed (2004) and Berry and Linoff (2004) also defined DM as the process of extracting or detecting hidden patterns or information from large databases. With an enormous amount of customer data, DM technology can provide business intelligence to generate new opportunities (Bortiz & Kennedy, 1995; Fletcher & Goss, 1993; Langley and Simon, 1995; Lau, Wong, Hui, & Pun, 2003; Salchenberger, Cinar, & Lash, 1992; Su, Hsu, & Tsai, 2002; Tam & Kiang, 1992; Zhang, Hu, Patuwo, & Indro, 1999).

Recently, a number of DM applications and prototypes have been developed for a variety of domains (Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro, & Simoudis, 1996) including marketing, banking, finance, manufacturing and health care. In addition, DM has also been applied to other types of data such as time-series, spatial, telecommunications, web, and multimedia data. In general, the DM process, and the DM technique and function to be applied depend very much on the application domain and the nature of the data available.

1.3. Relationship between KM and DM

Most KM and DM techniques involve learning patterns from existing data or information, and are therefore built upon the foundation of machine learning and artificial intelligence. The primary KM and DM techniques that can be used by the organizations include statistical analysis, pattern discovery and outcome prediction. A variety of non-typical data can be similarly monitored. Before the advent of DM and KM techniques, the organizations relied almost exclusively on human expertise. It was believed that these domain experts could effectively convert their collected data into usable knowledge. As the different types of data collected grew in scope, the organizations sought to find more practical methods to make sense of what they had. This led first to the employment of in-house statisticians who created better measures of performance and better decision-making criteria. One way that these measures were used was to augment the decision-making of domain experts with additional knowledge and provide them with a competitive advantage. Armed with this knowledge, it was not a far step for organizations to begin harnessing more practical methods of extracting knowledge using DM techniques. These techniques allowed organizations to begin to predict and/or forecast under specific conditions.

2. Material and methodology

2.1. Research material

Weingart (2003), Weingart (2004) pointed at the very influential role of the monopolist citation data producer ISI (Institute for Scientific Information, now Thomson Scientific) as its commercialization of these data (Adam 2002) rapidly increased the non-expert use of bibliometric analysis such as rankings. The materials used in this study were accessed from the database of the Social Science Citation Index (SSCI), obtained by subscription from the ISI, Web of Science, Philadelphia, PA, USA. In this study, we discuss the papers published in the period from 1989 to 2009 because there was no data prior to that year. The Social Sciences Citation Index is a multidisciplinary index to the journal article of the social sciences. It fully indexes over 1950 journals across 50 social sciences disciplines. It also indexes individually selected, relevant items from over 3300 of the world's leading scientific and technical journals.

2.2. Research methodology

Pritchard (1969, p. 349) defined bibliometrics as "the application of mathematics and statistical methods to books and other media of communication." Broadus (1987, p. 376) defined bibliometrics as "the quantitative study of physical published units, or of bibliographic units, or of the surrogates for either." Bibliometric techniques have been used primarily by information scientists to study the growth and distribution of the scientific article. Researchers may use bibliometric methods of evaluation to determine the influence of a single writer, for example, or to describe the relationship between two or more writers or works. Besides,