Data mining & Knowledge Management through Information Technology- Research Article

Prof. Tanvir Hussein
Associate Professor & Dean (Quality Education) Landmark Technical Campus, Amroha, UP, India
correspondence author: drtanvirhussein@gmail.com
***Salim Hyder Khan*
Research Scholar working on Knowledge Management from Shri Venketeshwara University.
Correspondence author : salimhyder@hotmail.com

Abstract
In recent years, with the advent, introduction and booms of information technology, businesses increasingly develop the capability to accumulate astronomical amounts of data in large databases. The availability of large volume of data, has made possible by new information technology tools, has created opportunities as well as challenges for businesses to leverage the data and gain competitive advantage. The present paper explores the systems and ways of taming the astronomical size data through information technology so as to make this data useful for the organization knowledge.

Key Words : Information technological Tool, Taming the astronomical data.

Introduction
The Internet and the World Wide Web have made the process of collecting data easier, adding to the volume of data available to businesses. On the one hand, many organizations have realized that the knowledge in these huge databases are key to supporting the various organizational decisions. Particularly, the knowledge about customers from these databases is critical for the marketing function. But, much of this useful knowledge is hidden and untapped. Under these conditions, data mining tools can help uncover the hidden knowledge and understand customer better, while a systematic knowledge management effort can channel the knowledge into effective marketing strategies. This makes the study of the knowledge extraction and management valuable.

The knowledge economy
In the new age of knowledge economy, knowledge has become a valuable resource and knowledge workers play a vital role (Drucker 1993). The defining characteristic of this age is the significance of knowledge workers and the importance of applying and developing new knowledge (Drucker 1993). Knowledge is applied to generate new knowledge in a continuous cycle (Castells 1996).

Knowledge is believed to be the major source of competitive advantage (Pan and Scarbrough 1998, Scarbrough et al. 1999) because of its "tacitness, inimitability and immobility (Grant 1997). "The capability to gather, lever and use knowledge effectively will become a major source of competitive advantage in many businesses over the next few years" (Trussler 1998). Thus, Knowledge Management (KM) is regarded as core competitive competence on which the success of organizations rely (Skyrme and Amidon 1998).

Knowledge Management and Knowledge
There is little consensus on the definition of KM (Bhatt 2001, Neef 1999), Pan and Scarbrough (1999) define KM as:

"The capacity (or processes) within an organization to maintain or improve organizational performance based on experience and knowledge." and It can be described as the way organizations build, supplement and organize knowledge and routines around their activities and within their culture and develop organizational efficiency by improving the use of employee skills.'

Effective KM requires an understanding of what constitutes knowledge (Allee 1997), because how it is managed depends on how it is viewed (Carlsson et al. 1996, Scarbrough et al. 1999).

There is also plenty of debate on the definition of knowledge. In the literature, knowledge is commonly categorized into distinctive types (McAdani and McCreedy 1999a, Venters 2002). In one of the most influential works on KM 'The knowledge creating company', Nonaka and Takeuchi (1995b) draw on Polanyi's (1967) work and develop a knowledge creation model based on two categories of knowledge - tacit and explicit knowledge. Tacit knowledge is personal, context-specific, and therefore hard to formalize and communicate. Explicit or 'codified' knowledge, on the other hand, refers to knowledge that is transmittable in formal, systematic language" (Nonaka and Takeuchi 1995).

Some authors define knowledge by distinguishing knowledge from information, and information from data in a hierarchical structure, each vary from another on certain characteristics, e.g. usefulness (Dretske 1981, Machlup 1980, Vance 1997). This approach is criticized by Alavi and Leidner (2001) as information and knowledge cannot be effectively distinguished on some dimensions e.g. "content, structure, accuracy, or utility".

Extending (Collins 1993) categorization of knowledge types, (Blackler 1995) has identified five types of knowledge implied in the literature: embedded, embodied, encultured, embrained and encoded. He criticizes that the taken-for-granted assumptions about the nature of knowledge which underpin conventional categorizations of knowledge offer a compartmentalized and static view of knowledge. He argues different types of knowledge should not be conceived as discrete and independent, and the concept of differentiating knowledge into categories should be abandoned. Similarly, Scarbrough et al. (1999) argue that the fixation on ontological debates about the nature of knowledge tends to promote prescriptive approaches as universal panaceas.

The functionalist and interpretive perspectives on KM

The debates in the KM literature could be broadly distinguished into two schools of thoughts based on the work of Schultze (1998). Schultze identifies two opposite perspectives on theories of KM: functionalist perspective and interpretive perspective.

Functionalist perspective sees knowledge as an objective representation of the world, existing in a number of forms and locations (Schultze 1998. Venters 2002). Whereas, interpretive perspective considers that knowledge only exists through human experience and social activities (SchuLtze 1998).

The emphasis on codification reflects the domination of technological approach to KM i.e. implementing new IT systems, databases, intranets, etc. as KM solutions (Scarbrough et al. 1999). The motivation behind this is to "stockpile workers' knowledge" so it is accessible to others (Cole-Gomolski 1997).

The functionalist discourse has been challenged on its attention on explicit knowledge and on the use of technology for codification (Spender 1996, Swan et al. 1999, Tsoukas 1996). The underlying assumption of the functionalist perspective is that all knowledge can be codified, which is not true (Gardner 1998,

Wenger et al. 2002) Interpretive proponents argue that knowledge cannot be managed like an object, separated from its human context (Choo 1996, Stenmark 2000-2001, Wenger et al 2002).

Increasingly, attention is focused on the influence of social structures and communities (e.g. communities of practice) on the management of knowledge and learning (Brown and Duguid 2001, Brown and Duguid 1991), exploring how tacit knowledge is embedded within social groups (Blackler 1995). However, it is more difficult to diffuse knowledge, as it requires a shared systems of meaning (Bresnen et al. 2003) that allow others to understand and apply knowledge in their context (Senge 1994. Spender 1996, Weick 1995). A community of practice is a social structure within which knowledge is created and shared (Brown 1998, Brown and Duguid 2000, Venger 1998, Wenger et al. 2002). It is defined as:

“A community of practice is a group of people who share a concern a set of problems, or a passion about a topic and who deepen their knowledge and expertise in this area by interacting on an ongoing basis" (Wenger et al. 2002).

Knowledge is perhaps best understood as multilayered and multifaceted, comprising cognition, action and resources, but also social relations” (Scarbrough et al 1999).

"Knowledge is multifaceted and complex, being both situated and abstract, implicit and explicit, distributed and individual, physical and mental, developing and static, verbal and encoded” (Blackler 1995).

Information Technology for knowledge management

Early debates on KM favour deployment of ICT (Cole-Goznolski 1997. Finerty 1997), hence the technology used in the industry has focused on capturing, codifying and reusing knowledge (Reiner and Fruchter 2000, Scherer and Reul 2000). However, recent KM studies in the industry are still emphasizing technological applications (Bakis and Sun 2000, Bouchlahern et al. 2000, Doherty 2000, Egbu 2000, Egbu et al. 2001, Fruchter et al. 2000, Moms et al. 2003, Robinson et al. 2001)

There are four classes of IT used for KM purposes (Laudon and Laudon 1998) in the industry.

- The first class concerns with knowledge creation e.g. CAD systems
- The second class belongs under office automation systems e.g. word processors, databases.
- The third class is systems that facilitate knowledge sharing e.g. Intranets, groupware, document management systems, bulletin boards, electronic mail etc.
- The fourth class is systems for knowledge capture and codification with AI technology (Carrillo et al. 2000).

Developments in database processing, data warehousing, machine learning and knowledge management have contributed greatly to our understanding of the data mining process. More recent research on data mining and knowledge discovery has further enhanced our understanding of the application of data mining and the knowledge discovery process. But, most research has focused on the theoretical and computational process of pattern discovery and a narrow set of applications such as fraud detection or risk prediction. The Data mining tasks is divided as follows

- *Dependency analysis*

- *Class identification*
- *Concept description*
- *Deviation detection*
- *Data Visualization*

Data mining tasks

Data mining is the process of searching and analyzing data in order to find implicit, but potentially useful, information. It involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and ultimately comprehensible information, from large databases.

Data mining uses a broad family of computational methods that include statistical analysis, decision trees, neural networks, rule induction and refinement, and graphic visualization. Although, data mining tools have been available for a long time, the advances in computer hardware and software, particularly exploratory tools like data visualization and neural networks, have made data mining more attractive and practical.

Dependency analysis

The primary type of dependency knowledge is the association between sets of items stated with some minimum specified confidence. This is also called "market basket analysis" and gives us the relationship between different products purchased by a customer. This type of knowledge can be useful in developing marketing strategies for promoting products that have dependency relationships in the minds of the customers.

Class identification

Class identification groups customers into classes, which are defined in advance. There are two types of class identification tasks — mathematical taxonomy and concept clustering. Mathematical taxonomy algorithms produce classes that maximize similarity within classes but minimize similarity between classes.

A drawback of this task is its inability to use background information, such as domain knowledge, to facilitate clustering. Concept clustering overcomes this limitation and determines clusters according to attribute similarity as well as conceptual cohesiveness as defined by domain knowledge. Users provide the domain knowledge by identifying useful clustering characteristics.

Concept description

Concept description can be used for summarization, discrimination, or comparison of marketing and customer knowledge. Data summarization is the process of deriving a characteristic summary of a data subset that is interesting with respect to domain knowledge and the full data file. Using summarization, a marketer can learn about customer characteristics by grouping them according to their occupation, income, spending patterns and types of purchases, and build customer profiles. Discrimination describes qualities sufficient to differentiate records of one class from another. Comparison describes the class in a way that facilitates comparison and analysis with other records. Comparison analysis can be done by statistical or visualization techniques.

Deviation detection

Deviations are useful for the discovery of anomaly and changes. Anomalies are things that are different from the normal. Anomalies can be detected by analysis of the means, standard deviations, and volatility measures from the data. In addition to anomalies, variables or attributes may have significantly

different values from the previous transactions for the same customer or group of customers. A credit card company may find a sudden increase in the credit purchases of an individual customer. This change in behavior can be a result of a change in the status of the customer, and not necessarily a fraud. Thus, confirmation of the "change" is made after investigation and the knowledge is updated.

Data Visualization

Data visualization software allows marketers to view complex patterns in their customer data as visual objects complete in three dimensions and colors.

They also provide advanced manipulation capabilities to slice, rotate or zoom the objects to provide varying levels of details of the patterns observed. To explore the knowledge in database, data visualization can be used alone or in association with other tasks such as dependency analysis, class identification, concept description and deviation detection. Keim (1996) provides an elaborate analysis of visualization techniques for mining large databases and classifies visualization techniques into pixel-oriented, geometric projection and graph-based. The pixel oriented technique maps each data value to a colored pixel and presents the data values belonging to each attribute in separate windows. Geometric projection techniques aim at finding interesting projections of multidimensional data set. The basic idea of the graph-based technique is to effectively present a large graph using specific layout algorithms, query languages, and abstraction techniques. Examples of graph based representations are 2-dimensional graphs, 3-dimensional graphs, Hygraphs and SeeNet.

Conclusion

Though data mining techniques are used in several areas such as fraud detection, bankruptcy prediction, medical diagnosis, and scientific discoveries, their use for marketing decision support highlights unique and interesting issues such as customer relationship management, real-time interactive marketing, customer profiling and cross-organizational management of knowledge.

In the current customer centric business environment, it is our firm belief that there is a need for deeper understanding of use of data mining and knowledge management for marketing decision support. Towards that end, in this paper, we have shown how large volume of data can be organize using IT techniques of data mining and can utilized data as organizational knowledge management framework.

With the availability of large volume of data, made possible by modern information technology, a major problem is to filter, sort, process, analyze and manage this data in order to extract the information relevant to the user. The growth in the size and number of existing databases far exceeds human abilities to analyze such data using traditional tools and thus creates both a need and an opportunity for data mining tools. A systematic application of data mining techniques will enhance the knowledge management process and arm the companies with better knowledge of their customers leading to better service to them. To us, it is also clear that the Web technology will have a major impact on the practice of data mining and knowledge management and that should present interesting challenges for future information systems research.

References

- Allee (1997) "12 Principles of Knowledge Management'. Training and development, 51(11), pp. 71-74.
- Bakis. N. and M. Sun (2000) "Intelligent Borker for Collaborative Search and Retrieval of Construction Information on the www" in Proceedings of the International Conference on Construction Information Technology, Reykjavik. Iceland: Icelandic Building Research Institute.
- Blackler. F. (1995) "Knowledge, Knowledge Work and Organizations: An Overview and Interpretation", Organization Studies, 16 (6), p. 1021.
- Bhatt, G. D. (2001) "Knowledge Management in Organizations: Examining the Interaction between Technologies, Techniques, and People", Journal of Knowledge Management, 5 (1), p. 68±75.
- Brown, J. S and P. Duguid (2001) "Knowledge and Organization: A Social-Practice Perspective". Organization science, 12 (2), pp. 198-211
- Brown, S. J. and P. Duguid (1991) "Organizational Learning and Communities of Practice: Toward a Unified View of Working. Learning and Innovation", Organization Science, 2 (1), pp. 40-57.
- Carlsson, S. A. O. A. El Sawy. I. Eriksson and A. Raven (1996) "Gaming Competitive Advantage through Shared Knowledge Creation: In Search of a New Design Theory for Strategic Information Systems" in Proceedings of the fourth European conference on information systems on information systems, Lisbon, 1996.
- Castells, M. (1996) The Rise of the Network Society, Oxford: Blackwell.
- Collins, W (1993) "The Structure of Knowledge", social research, 60 pp, 95-116.
- Cole-Gomoiski, B. (1997) "Users Loathe to Share Their Know-How", Computerworld, 31(46), p. 6.
- Choo. C W. (1996) "The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions", International Journal of Information Management, 6 (5). pp. 329-340
- Demarest. M (1997) "Understanding Knowledge Management", Long Range Planning, 30 (3), pp. 374-384.
- Dretske. F. (1981) Knowledge and the Flow of Information, MIT Press. Cambridge. MA.
- Drucker, P. (1993) Post-Capitalist Society Oxford: Burtenvorth-Heinemann.
- Egbu. C. (2000) "The Role of Information Technology in Strategic Knowledge Management and Its Potential in the Construction Industry". in Proceedings of the UK National conference on Objects and integration for architecture, engineering and construction, Sponsor: Building Research Establishment Ltd., pp. 106-114.
- Egbu. C., C. Gaskell and J. Howes (2001) "The Role of Organisational Culture and Motivation in the Elective Utilisation of Information Technology for Teamworking in Salford. p. 91—100.
- Finerty. L. (1997) "Information Retrieval for Intranets: The Case for Knowledge Management", Document world, 2 (5). pp. 2-34.
- Fruchter. R., K. Reiner. S. Yen and A. Retik (2000) "Kiss: Knowledge and Information Slider System" in Proceedings of the International Conference on Construction Information Technology, Reykjavik. Iceland: Icelandic Building Research Institute.
- Gardner. D. (1998) "Knowledge That Won't Fit the Database - People". InfoWorld, 20 (14), p. 98.
- Grant. R. (1997) "The Knowledge-Based View of the Firm: Implications for Management Practice". Long Range Planning, 30 (3), pp. 450-454.
- Kakölä, T. K. (1995) "Increasing the Interpretive Flexibility of Information Systems through Embedded Application Systems", Accounting, management, and information technologies 5 (I), pp.79-102.

- Keim, D.A. (1996) Visualization techniques for mining large databases: a comparison, IEEE Transactions on Knowledge and Data Engineering 8 (6) pp.923–937.
- Laudon. K. C. and P. L. Laudon (1998) Management Information Systems, Prentice-Hall, New Jersey. USA.
- Machlup. F. (1980) Knowledge: Its Creation, Distribution, and Economic Significance, Princeton University Press. Princeton. NJ.
- McAdam. R. and S. McCreedy (1999b) "The Process of Knowledge Management within Organizations: A Critical Assessment of Both Theory and Practice', Knowledge and process management, 6(2), pp. 101-112.
- M.J. Shaw, Machine learning methods for intelligent decision support: an introduction, Decision Support Systems 10 (2), 1993 79–83.
- Neef. D. (1999) "Making the Case for Knowledge Management: The Bigger Picture". Management Decision, 37 (1), p. 72.
- Nonaka. I. and H. Takeuchi (1995), The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation.
- Pan. S. L. and H. Scarbrough (1998) "A Socio-Technical View of Knowledge-Sharing at Buckman Laboratories". Journal of Knowledge Management, 2 (1), p. 55.
- Pan. S. L. and H. Scarbrough (1999) "Knowledge Management in Practice: An Exploratory
- Case Study". Technology analysis & Strategic management, 11(3).
- Pentland. W T. (1995) "Information Systems and Organizational Learning: The Social Epistemology of Organizational Knowledge Systems". Accounting, Management and Information Technologies, 5 (1), PP. 1-21.
- Robey, D and S. Sahay (1996) "Transforming Work through Information Technology: A Comparative Case Study of Gis in County Government", Information Systems Research, 7(I), pp. 93-110.
- Robey, D., N. A. Wishart and K G. Rodnguez-Diaz (1995) "Merging the Metaphors for Organizational Improvement: Business Process Reengineering as a Component of Organizational Learning". Accounting, management, and information technologies, 5(I), pp. 23-39
- Robinson. H. S. P M. Carrillo. C. 3. Anumba and A. M. Al-Ghassani (2001) "Linking Knowledge Management Strategy to Business Performance in Construction Organisations", in Proceedings of ARCOM 17th Annual Conference, University of Salford. p. 577—586.
- Scarbrough. H., J. Swan and J. Preston (1999) "Knowledge Management: A Literature Review" Institute of Personnel and Development.
- Scott, J. E. (2000) "Facilitating Organizational Learning with Information Technology". Journal of Management information systems, 17 (2), pp. 81-113.
- Schultze. U. (1998) "Investigating the Contradictions in Knowledge Management" in IFIP WG8.2 & WG8.6 Joint Working Conferences on Information Systems: Current Issues and Future Changes, Helsinki. Finland. Omnipress. Wisconsin. USA.
- Schultze. U. and D. E. Leidner (2002) "Studying Knowledge Management in Information Systems Research: Discourses and Theoretical Assumptions", MIS Quarterly 26 (3), pp. 213-242.
- Scherer. R. J. and S. Reul (2000) "Retrieval of Project Knowledge from Heterogeneous Documents" in Proceedings of the 8th International Conference on computing in civil and building engineering, Reston. VA: American Society of civil engineers. pp. 812-19.
- Senge. P M (1994) The Fijìh Discipline Fieldbook: Strategies and Tools for Building a Learning Organization, New York, Currency Doubleday.

- Skyrme, D. J. and D. M. Amidon (1998) 'New Measures of Success", The journal of business strategy, 19 (1), pp. 20-24.
- Spender, J.-C. (1996) "Organizational Knowledge, Learning and Memory: Three Concept in Search of a Theory", Journal of Organizational Change Management. 9 (I), pp. 63-78.
- Stenmark, D. (2000-2001) "Leveraging Tacit Organizational Knowledge". Journal of Management information systems, 17 (3), pp 9-24.
- Swan. J., S. Newell, H. Scarbrough and D. Hilop (1999) "Knowledge Management and Innovation Networks and Networking". Journal of Knowledge Management, 3 (4), pp. 262-275.
- Tsoukas, H. (1996) "The Firm at a Distributed Knowledge System A Constructionist Approach", Strategic Management Journal, 17 (Special Issue: Knowledge and the Firm), pp. 11-25.
- Vance. D. M. (1997) "Information. Knowledge, and Wir4om: The Epistemic Hierarchy and Computer-Based Information System" in Proceeding of the third Americans Conference on Information System, Indianapolis, IN.
- Venters, W. (2002) "Literature Review for C-Sand Knowledge Management" London School of Economics.
- Weick. K. (1995) Sen,remakmg ¡n Organization:. Sage. Thousand Oak,(CA).
- Wenger, E. (1998) Communities of Practice: Learning, Meaning and Identity, Cambridge University Press.
- Wenger. E., R. McDermott and W. Snyder (2002) Cultivating Communities of Practice, Harvard Business School Press, Boston, Massachusetts.

****

***Prof. Tanvir Hussein, Ph.D.**

Associate Professor & Dean (Quality Education) Landmark Technical Campus, Amroha, UP, India
Research Guide, Shri Venketeshwara University, wrote various articles on corporate governance & corporate social responsibility, worked in Kingdom of Saudi Arabia as Assistant Professor & Quality Education Head before repatriate. His main area of interest & work is Emotional Intelligence, Knowledge Management, CSR and Corporate Governance.
Can be reach at: drtanvirhussein@gmail.com

****Salim Hyder Khan**

Research Scholar working on Knowledge Management from Shri Venketeshwara University.
Can be reach at: salimhyder@hotmail.com

