

December 2003

Investigation of Information Retrieval Accuracy from Knowledge Management Systems: A Multi- Method Approach

Ryan LaBrie
Arizona State University

Follow this and additional works at: <http://aisel.aisnet.org/amcis2003>

Recommended Citation

LaBrie, Ryan, "Investigation of Information Retrieval Accuracy from Knowledge Management Systems: A Multi-Method Approach" (2003). *AMCIS 2003 Proceedings*. 446.
<http://aisel.aisnet.org/amcis2003/446>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

INVESTIGATION OF INFORMATION RETRIEVAL ACCURACY FROM KNOWLEDGE MANAGEMENT SYSTEMS: A MULTI-METHOD APPROACH

Ryan C. LaBrie
W. P. Carey School of Business
Arizona State University
ryan.labrie@asu.edu

Abstract

This extended abstract facilitates a discussion on the issues entailing the accuracy of information retrieval (IR) from knowledge management systems (KMS). This dissertation will be taking a multi-method approach in its investigation. First, an empirical examination of keyword usage from a representative IS research journal is performed to validate the need for this research. Second, a case study is underway with the Intel Corporation, on their satisfaction with the quality of their information retrieval capabilities. Lastly, a laboratory experiment is proposed to test the informational retrieval accuracy of two types of knowledge management system infrastructure – traditional relational-based and multidimensional-based KMS.

Theories developed in cognitive psychology suggest that the hierarchical nature of multidimensional systems, through its aggregation capabilities and tree-like structures, generate more accurate information, faster, and provide more closely related information to the knowledge worker. This leads to directional hypotheses that can be tested in an experiment. The results of these hypotheses have implications for knowledge management systems design. It is posited that KMS that are properly designed with more accurate information retrieval mechanisms (i.e. multidimensional technologies) will reduce organizational costs, in time and effort, by decreasing the wasted time spent investigating information that does not support the current knowledge requirements of the user.

Keywords: Accuracy, information retrieval, keyword limitations, knowledge management systems, KMS, methodology, multidimensional technologies

Introduction

Research conducted under the auspices of knowledge management varies greatly in direction and scope. In fact in recent years entire issues of leading MIS research journals have been dedicated to the topic of knowledge management and knowledge management systems or similar titles – JMIS (18:1) 2001, MISQ (26:3) 2002 and again forthcoming in 2004. Alavi & Leidner (2001) provide a foundational framework to help contextualize the various components of knowledge management systems research and development. In this framework they offer four focus areas for research; knowledge creation, knowledge storage/retrieval, knowledge transfer, and knowledge application. This dissertation work, while recognizing the importance of all four areas of knowledge management research, focuses specifically on the retrieval aspects from knowledge management systems.

We've all heard the acronym GIGO – garbage in, garbage out, and there has been a significant amount of research that has focused on building information systems to decrease the garbage in factor. This research, however, focuses on the problem where you might have quality information captured in a knowledge management system, but do not have an effective mechanism to retrieve that knowledge, and thus you are faced with the dilemma of quality in, but still garbage out.

Markus (2001) suggests a “Theory of Knowledge Reuse” in which the information system that supports the knowledge management initiative should handle both access to expertise and access to the experts. This allows for optimally managing both the explicit knowledge of the organization, the knowledge that is captured in the system as expertise, as well as an attempt to manage the tacit knowledge, which resides within the individuals (experts) of the organization. While this dissertation research focuses primarily on explicit knowledge that has already been captured and codified, its implications extend beyond expertise and apply to knowledge management of experts as well. In developing this theory of knowledge reuse Markus (2001) stresses that information technology plays an intermediary role. Specifically, Markus places greater emphasis on the creation and utilization of the repository. This dissertation echoes that emphasis and suggests that by re-evaluating the underlying infrastructure of the repository, the information retrieval process may be more effective.

Previous Research

Knowledge management research is a very broad area within the information systems discipline. Several authors (Alavi & Leidner 1999, 2001; Spiegler, 2000; Schultze & Leidner, 2002) have suggested various classification models that break down knowledge management into different domains. Alavi and Leidner (1999) surveyed 109 executives, obtaining 50 usable responses, on their perceptions of KMS activity within their firms and its potential benefits. This research identified three perspectives for knowledge management – an Information-based, a Technology-based, and a Cultural-based perspective. Among the conclusions from this research were: 1) knowledge management systems are multi-faceted; and 2) it is important to try to develop metrics to assess the benefits of KMS. In their 2001 work Alavi & Leidner (2001) propose a conceptual foundation that includes the knowledge management systems domains of knowledge creation, knowledge storage/retrieval, knowledge transfer, and knowledge application. Strategic research questions posed by Alavi and Leidner (2001) pertaining to knowledge retrieval include: 1) Is stored knowledge accessed and applied by individuals who do not know the originator of the knowledge? and 2) what retrieval mechanisms are most effective in enabling knowledge retrieval?

Spiegler (2000) concludes that knowledge management is indeed a new idea, rather than a recycled concept. The basis of his argument is that yesterday's data are today's information, and tomorrow's knowledge, which in turn, recycles back through the value chain of data-information-knowledge, represents another framework in which to investigate knowledge management.

Schultze and Leidner (2002) propose using Deetz's four discourses of organizational inquiry to classify IS knowledge management research. These categories include: normative, interpretive, critical, and dialogic discourse. Based on an analysis of the literature they find that the bulk of the research falls within the normative discourse classification, interpretive discourse is used less frequently, and lagging relatively far behind is KM research following a critical or dialogic discourse.

Each of the articles discussed above suggests various frameworks or classifications for knowledge management research within the information systems field. Table 1 summarizes their contributions.

Table 1. Knowledge Management Research Classifications

Author (Year)	KM Classification Context	Research Method
Alavi & Leidner (1999)	KM Perspectives: Information-based, Technology-based, and Culture-based	Executive Survey (109 Surveyed, 50 usable responses)
Alavi & Leidner (2001)	KM Processes: Creation, Storage/Retrieval, Transfer, and Application	Literature Review
Spiegler (2000)	KM Transformation: Data Processing, Information Processing, Knowledge Processing, and a reverse process	Literature Review Model Development
Schultze & Leidner (2002)	KM Discourses Normative, Interpretive, Critical, and Dialogic	Literature Review

Next, document management literature is examined. Document management systems have existed for a number of years and can arguably be considered some of the earliest knowledge management systems. Sprague (1995) describes how information systems

managers, if properly prepared, can take the next step beyond managing text and numbers to managing electronic document objects. These objects may take the form of contracts, email/voicemail, video clips, meeting transcripts, drawings/blueprints/photographs, or any number of object types. The contributions of this research include the idea that managing knowledge objects is different than managing basic text and numbers. Additionally, Gordon and Moore (1999) develop a foundation for a “reading system” that examines how a document is used and the purpose for which it is used. This reading system is a new type of information system developed with a formal language to help knowledge workers retrieve knowledge documents in a more effective manner.

An additional set of literature that is examined pertains to codification/classification systems and keyword usage (the term keyword is used in this abstract to mean both keyword and keyword phrases). This literature is important with respect to the first component of this dissertation work, the empirical investigation of keyword search limitations. Keywords play an important role in information retrieval, yet they have their shortcomings as well. Often to overcome these limitations classification categories are developed. Perhaps the most widely recognized work on keyword classification systems within the IS community is the work of Barki et al (1988, 1993). They have developed a classification system commonly referred to as the ISRL (Information Systems Research Literature) categories. These categories were developed based on keyword usage in top IS journals and are currently being utilized by leading IS journals including *MIS Quarterly*.

Self examination of a field via analysis of its publications is a common practice among researchers (Neufeld et al., 2002; Vessey et al., 2002; Swanson & Ramiller, 1993; Gorla and Walker, 1998). Vessey et al. (2002) analyzed the diversity within the IS discipline and its journals. They produced a five part classification scheme based on the following categories: reference discipline, level of analysis, topic, research approach, and research method. Similarly, Swanson and Ramiller's (1993) paper on information systems research thematics analyzed submissions to a new journal to discover themes and relationships among IS research. Neufeld et al. (2002) explored the relationship of IS topics published in non-IS business disciplines (such as accounting, marketing, etc.).

Gorla and Walker (1998) suggest that searches are not effective unless an unambiguous keyword list is universally accepted. They collected their data from ABI/Inform database for the top MIS journals over the 10 year period from 1984-1994. Their analysis was performed on 14,676 articles, 3305 keywords, and 121,548 occurrences of those keywords. Similarly, the first study within the scope of this dissertation also intends to shed light on the limitations of keywords. It extends the discussion to include a demonstration that classification schemes can overcome some of these limitations. It then also suggests an infrastructure framework that may be more conducive to information retrieval from knowledge management systems.

The rest of this extended abstract proceeds as follows: the next section provides the overall research question and develops supporting hypotheses. The following section introduces the methodological approaches that are to be taken during this dissertation research. Several research methods will be employed including: 1) an empirical investigation to the limitations of keywords in knowledge management systems, 2) a field (case) study that includes qualitative discussion as well as a survey component questioning an organization's satisfaction with its current retrieval capabilities, and 3) a laboratory experiment which will examine two KMS infrastructures to determine if there is a superior retrieval method. Next a discussion section presents findings from the first study and provides some comments on expected results. The final section provides summary remarks including avenues for future research. The paper concludes with a list of references.

Research Question and Hypotheses

This research focuses specifically on the information retrieval aspects of knowledge management systems. Within that defined scope the following research question is presented and investigated: *To what extent do knowledge management systems provide accurate information retrieval?*

Based on this research question the following preliminary hypotheses have been generated to investigate whether current knowledge management systems are employing the most effective means for accurate information retrieval.

Hypothesis 1: Knowledge management systems that employ multidimensional technology will produce more accurate results than knowledge management systems that employ relational technology.

Or more formally:

Hypothesis 1a: Multidimensional KMS will produce fewer Type I errors (false-positives) than relational KMS.

Hypothesis 1b: Multidimensional KMS will produce fewer Type II errors (false-negatives) than relational KMS.

The proposition behind hypothesis 1 can best be described in the following context. Suppose you are an IS researcher and have a document management system that contains various research articles. You are doing research on the Technology Acceptance Model (TAM) and you would like to find all the articles that have tested TAM. Presume that the system actually had 10 articles that truly tested TAM. Through your querying of the system you discover 9 articles. However, upon closer inspection of those articles only seven of them truly tested TAM. You have found only 70% of all the relevant articles. Furthermore, you have produced 2 false positives, which waste your time by not providing the information you requested, and 3 false negatives, or 3 articles that you should have found, but did not find.

Hypothesis 2: Knowledge management systems that employ multidimensional technology will enable users to find the intended information faster than knowledge management systems that employ relational technology.

A knowledge worker, given enough time, may be able to find all 10 of the relevant TAM articles by running multiple queries. Hypothesis 2 states that using a multidimensional system will enable the knowledge worker to discover the articles faster than using a relational system.

Hypothesis 3: Knowledge management systems that employ multidimensional technology will enable users to discover more (relevant/closely) related information than knowledge management systems that employ relational technology.

Cognitive psychology research has extensively studied information retention and retrieval. In particular the debate of recall (keyword) versus recognition (hierarchy) has long been studied. Results have found that recognition is always superior (Anderson, 1995, Bower et al. 1969). In fact, Anderson concludes that, “retrieval of information is facilitated if it is organized hierarchically” (Anderson, 1995 p.223).

Proposed Research Method

This dissertation research employs a multi-method approach in investigating the issue of information retrieval accuracy from knowledge management systems. Multi-method research approaches have been extolled in the IS literature by leading researchers (Kerlinger & Lee, 2000; Yin, 1989 & 1994; Kaplan & Durchon, 1988). By engaging in a multi-method approach the findings from this stream of research form a more complete picture, strengthening the findings and support or non-support of the hypotheses presented.

In the first phase of this dissertation research, data was collected and analyzed to determine if there was an accuracy issue with current knowledge management systems retrieval mechanisms. A common practice for most KMS is to rely on keyword phrases as the primary mechanism for search and retrieval. While this has suited knowledge workers in the past, it is not without its limitations. To demonstrate these keyword search limitations a database was developed to analyze all of the articles that have appeared in *MIS Quarterly* during its 27 year history. The dataset includes 603 articles with 1770 keywords occurring 2857 times, and 610 classification categories occurring 2280 times. Results can be found in LaBrie & St. Louis (2003).

In part two of this research a field (case) study is under investigation in conjunction with the Intel Corporation and their EPIK (Enabling People through Innovation and Knowledge). Along with the rich set of qualitative case study notes, a survey is being developed to measure their knowledge workers’ satisfaction with the current knowledge management information retrieval capabilities.

The final component of this dissertation research is the development of a laboratory experiment to test the hypotheses presented above. Two IS research literature knowledge management systems will be developed, one using a relational infrastructure, and one using a multidimensional infrastructure. Knowledge workers in the form of faculty, graduate students, and undergraduate students will be tested on one or both of the systems to determine the more effective system based on accuracy, speed, and amount of related information. The general linear model statistical method will then be employed to determine if the results are statistically significant.

Discussion (Including Preliminary Findings)

The findings from an initial investigation (LaBrie & St. Louis, 2003) of information retrieval from knowledge management systems employing keyword search mechanisms suggest that very inaccurate results can occur. Two findings of note from this research included the lack of keyword reuse and the large number of ambiguous and redundant terms. First, the findings indicate that 76% of all keywords employed were used only once, and a mere 10% were used more than twice. This caused the authors to question what makes a good or bad keyword. Our conclusions suggest good keywords are common (i.e. have generally been used before), unambiguous, utilize short phrase (i.e. are not 10 words in length) and do not employ plurals, the word “and,” hyphens, slashes, or parentheses except in rare circumstances. Second, with respect to ambiguity this study provided, as just one example, 14 different variations of the keyword phrase “System Design.” The study then went on to examine the value of classification schemes in overcoming some of the issues surrounding keywords, and suggests a multidimensional infrastructure to support knowledge retrieval.

The second component of this research involves a field (case) study. The Intel Corporation has agreed to an assessment of their current information retrieval practices. Along with the qualitative interviews and field notes, a survey is being developed which will be administered to their knowledge workers for assessing their satisfaction with their current (keyword search-based) knowledge management systems.

Finally progress is being made on the development of a laboratory experiment that will test the hypotheses presented above. The context (a document management system) has been selected and the initial relational system infrastructure has been developed. Next, a multidimensional infrastructure is now in the design phase. Upon completing the development of the multidimensional KMS, an appropriate experimental design will be developed in order to test the hypotheses.

Conclusion

Knowledge management research within the information systems discipline is a broad domain that contains many facets for exploration. This research focuses on the information retrieval aspect, within the knowledge creation-storage/retrieval-transfer-application framework developed by Alavi & Leidner (2001). Theoretical foundations come from Markus’ (2001) “theory of knowledge reuse” and the cognitive psychology reference discipline, which suggests that perhaps there is a more effective way for information retrieval from knowledge management systems.

The current status of this research, as of mid-March 2003, is a work-in-progress toward a dissertation proposal. The proposal defense is expected to occur in May 2003. The results from the first phase study on keyword search limitations can be found in LaBrie & St. Louis (2003). The field study phase currently is in progress and is producing lots of qualitative data. The knowledge user survey is also expected to lend support to this stream of research. Finally, one of two knowledge management systems has been developed for the laboratory experiment.

Upon the successful completion of this dissertation research it is expected that the findings will support the conjecture that knowledge management systems employing multidimensional technologies provide superior information retrieval vis-à-vis traditional relational technologies.

Future Research: Further research will be necessary to generalize these findings to additional contexts. Future research may include testing the experiment within a different environment. Another context for this experiment to be replicated in is within the corporation, using areas of expertise as the knowledge object. Discussions are under way with Intel to implement a multidimensional knowledge management system, pending the finding from the survey. Additionally, further development of a model or framework suggesting how to best implement multidimensional KMS for accurate information retrieval might prove to be fruitful. Finally, research investigating tying actual cost (ROI) to the accuracy (i.e. Type I and II errors) produced by inferior systems would benefit both academics and practitioners in the study of knowledge management systems.

References

Alavi, M., and Leidner, D.E. “Knowledge Management Systems: Issues, Challenges, and Benefits,” *Communications of the Association for Information Systems*, (1:7), 1999, pp. 1-37.

- Alavi, M., and Leidner, D.E. "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS Quarterly*, (25:1), March 2001, pp. 107-136.
- Anderson, J.R. *Cognitive Psychology and its Implications*, 4th Ed., New York, NY: W. H. Freeman and Company, 1995.
- Barki, H., Rivard, S., and Talbot, J. "An Information Systems Keyword Classification Scheme" *MIS Quarterly*, (12:2), June 1988, pp. 299-323.
- Barki, H., Rivard, S., and Talbot, J. "A Keyword Classification Scheme for IS Research Literature: An Update," *MIS Quarterly*, (17:2), June 1993, pp. 209-226.
- Bower, G.H., Clark, J.C., Lesgold, A.M., and Winzenz, D. "Hierarchical Retrieval Schemes in Recall of Categorical Word Lists," *Journal of Verbal Learning and Verbal Behavior*, (8), 1969, pp. 323-343.
- Davis, F.D. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, (13:3), September 1989, pp. 319-340.
- Gordon, M.D., and Moore, S.A. "Depicting the use and purpose of documents to improve information retrieval," *Information Systems Research*, (10:1), March 1999, pp. 23-37.
- Gorla, N. and Walker, G. "Is The Lack of Keyword Synergism Inhibiting Maturation in the MIS Theory? An Exploratory Study," *Information Processing & Management*, (34:2), 1998, pp. 325-339.
- Kaplan, B. and Durchon, D. "Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study," *MIS Quarterly*, (12:4), December 1988, pp. 571-586.
- Kerlinger, F. and Lee, H. *Foundations of Behavioral Research*, 4th Ed., Fort Worth, TX: Harcourt College Publishers, 2000.
- LaBrie, R.C. and St. Louis, R.D. "Information Retrieval from Knowledge Management Systems: Using Knowledge Hierarchies to Overcome Keyword Limitations," *Ninth Americas Conference on Information Systems*, Tampa, FL, Aug. 4-6, 2003.
- Markus, M.L. "Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success," *Journal of Management Information Systems*, (18:1), Summer 2001, pp. 57-93.
- Schultze, U. and Leidner, D.E. "Studying knowledge management in information systems research: Discourses and theoretical assumptions," *MIS Quarterly*, (26:3), September 2002, pp. 213-242.
- Sprague, R.H. "Electronic Document Management - Challenges and Opportunities for Information-Systems Managers," *MIS Quarterly*, (19:1), March 1995, pp 29-49.
- Spiegler, I. "Knowledge Management: A New Idea or a Recycled Concept?" *Communications of the Association for Information Systems*, (3:14), 2000, pp. 1-23.
- Swanson, E.B. and Ramiller, N.C. "Information Systems Research Thematics: Submissions to a New Journal, 1987-1992," *Information Systems Research*, (4:4), December 1993, pp. 299-330.
- Vessey, I., Ramesh, V., and Glass, R.L. "Research in information systems: An empirical study of diversity in the discipline and its journals," *Journal of Management Information Systems*, (19:2), Fall 2002, pp 129-174.
- Yin, R.K. "Research Design Issues in Using the Case Study Method to Study Management Information Systems." In J.I. Cash, Jr. and P.R. Lawrence (Eds.) *The Information Systems Research Challenge Vol. 1 - Qualitative Research Methods*, Cambridge, MA: Harvard Business School Publishing, 1989.
- Yin, R. *Case Study Research: Design and Methods*, 2nd Ed., Thousand Oaks, CA: SAGE Publications, 1994.