December 2003

Information Retrieval from Knowledge Management Systems: Using Knowledge Hierarchies to Overcome Keyword Limitations

Ryan LaBrie
Arizona State University

Robert St. Louis
Arizona State University

INFORMATION RETRIEVAL FROM KNOWLEDGE MANAGEMENT SYSTEMS: USING KNOWLEDGE HIERARCHIES TO OVERCOME KEYWORD LIMITATIONS

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

Robert St. Louis
W. P. Carey School of Business
Arizona State University
st.louis@asu.edu

Abstract

Information retrieval (IR) is a key component of knowledge management systems (KMSs). KMSs frequently rely on keyword searches as a primary mechanism for retrieval. While keyword searches are very helpful to knowledge workers, they have their limitations. To illustrate some of the limitations of keyword searches, a database containing all of the articles that have appeared in MIS Quarterly *was constructed. Prior research shows that ambiguous, poorly constructed keyword phrases lead to poor information retrieval results. Knowledge hierarchies have been used to overcome some of the keyword searches limitations. This research proposes the use of dimensional modeling and multidimensional database technologies to implement knowledge hierarchies. It suggests that many of the limitations inherent in keyword searches can be eliminated from knowledge management systems by exploiting the benefits of the hierarchical structure that underlies multidimensional databases.*

Keywords: Classification, information retrieval, keywords, knowledge management systems, KMS, multidimensional databases, searching

Introduction

Research conducted under the auspices of knowledge management varies greatly in direction and scope. 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. To help contextualize the various components of knowledge management systems research, Alavi & Leidner (2001) provide a foundational framework for knowledge management research and development. In this framework they offer four focus areas for research: knowledge creation, knowledge storage/retrieval, knowledge transfer, and knowledge application. This paper, while recognizing the importance of all four areas of knowledge management research, focuses specifically on the retrieval aspects of knowledge management systems.

We've all heard the acronym GIGO – garbage in, garbage out. A significant amount of research has focused on building information systems to decrease the garbage in factor. It is possible, however, to have quality information captured in a knowledge management system, but to have no effective mechanism to retrieve that knowledge. Thus you are faced with the dilemma of quality in, but still garbage out.

Markus (2001) presents a "Theory of Knowledge Reuse" in which the information system that supports the knowledge management initiative handles 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 the tacit knowledge of the organization (which resides within the individual experts). While this research focuses primarily on explicit knowledge

that has been captured and codified, its implications extend beyond expertise and apply to knowledge management of experts as well. In developing her theory of knowledge reuse Markus (2001) argues that information technology plays an intermediary role in knowledge management. Specifically, she emphasizes the creation and utilization of the repository. This study echoes that emphasis and suggests that by re-evaluating the underlying infrastructure of the repository, the effectiveness of the information retrieval process may be increased.

This paper is organized as follows: the next section presents prior literature on knowledge management. Specifically the literature review incorporates discussion on classifications of knowledge management systems research, on document management, and on keyword searches. The following section presents the research questions and method used in examining keyword limitations, followed by a brief analysis of the data. A discussion section is then presented that suggests an infrastructure that might alleviate some of these keyword limitations. The final section provides concluding remarks including limitations to this research and avenues for future research.

Prior Literature

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 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 KMSs. 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 that 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. This represents yet 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, followed by interpretive discourse, and lagging relatively far behind KM research following a critical or dialogic discourse.

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

Table 1. Knowledge Management Classification Literature

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, the 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 "readying system" that examines how a document is used and the purpose for which it is used. This readying 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 paper to mean both keyword and keyword phrases). This literature is important with respect to the empirical investigation of keyword search limitations. Keywords play an important role in information retrieval, yet they have their shortcomings. Often times, 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 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 the 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 analysis of the *MISQ* database in this paper sheds light on the limitations of keywords. It first extends the discussion to include a demonstration that classification schemes can overcome some of these limitations; and then suggests an infrastructure framework that may be more conducive to information retrieval from knowledge management systems.

Method

In this paper we follow the spirit of the Gorla and Walker paper by evaluating a unique dataset that the authors have constructed. Our data set is assembled from the keywords for articles that appeared in *MIS Quarterly*. Data was collected in two fashions. For the most recent articles (1994-present), it was collected electronically from the *MIS Quarterly* Internet website (**www.misq.org**). All data from 1977 – Volume 1, Issue 1, to 1994 was entered by hand by one of the authors, gathering the keywords from the original text. The purpose of this data set was to answer the following research questions.

- How diverse is the use of keywords in information systems research?
- How widespread are keywords that duplicate meaning or are ambiguous, such that they may cause problems with keyword search results?
- Is there a correlation between the use of keywords and classification categories?

The first research question is similar to Vessey et al.'s research question 1b, stated by the authors as, "How diverse are IS Journals in the topics they publish?" (Vessey et al., 2002 pp.135). They developed a five-part classification system, and then hand categorized each article, from five journals over a five year period. In this research we choose to examine one representative journal over its 27 year history of existence.

The second research question is similar to Gorla and Walker's question, "Are there keywords that have essentially the same meaning that could be replaced with less ambiguous keywords to make a keyword search more effective and less time consuming?" (Gorla & Walker, 1998 pp.328). Gorla and Walker refined their keyword list down to 1056 keywords from an initial

list of 4386, or an approximate 75% reduction. This refinement process was performed so as to compare their list to the Barki et al. (1988, 1993) ISRL categories. Our purpose for investigating this question is to focus on the inherent limitations of basic keyword search mechanisms available in most KMSs, and to show viable alternatives.

The third question we pose focuses on the similarities and differences between keywords and classification categories. Is there a one-to-one mapping of keywords to categories, or are these two retrieval mechanisms disjoint? Furthermore, we propose an investigation on whether or not one is superior to the other.

Analysis

In all 608 articles (minus 20 without any keywords) were examined, with 1791 unique keywords/keyword phrases and 2885 occurrences of those keywords found. One of the first interesting findings was the lack of reuse of keywords. Inherently one would assume that keywords provide for a commonality among articles. That is one would suspect that a common keyword term, such as "case study", "system design", or "database" would be used by a number of different articles relating to those terms. In fact the analysis presents quite a different picture. Table 2 shows that the vast majority of keywords used in *MIS Quarterly* articles (77%) are used only once, and that a mere 10% of the keywords are used more than twice.

Table 2. Frequency of Keyword Usage

Number of times a keyword was used		Percent of Total
1	1375	76.77%
2	234	13.07%
>2	182	10.16%
Total	1791	100.00%

Appendix A presents a list of the top keywords used in *MIS Quarterly*. To qualify for this listing a keyword phrase had to be used in at least ten different articles.

Table 3 presents some basic descriptive statistics on the keyword data, while Figure 1 charts the frequency distribution of keywords per article.

Table 3. Descriptive Statistics of Keywords per Article

4.90	mean
4	median
4	mode
2.13	stdev

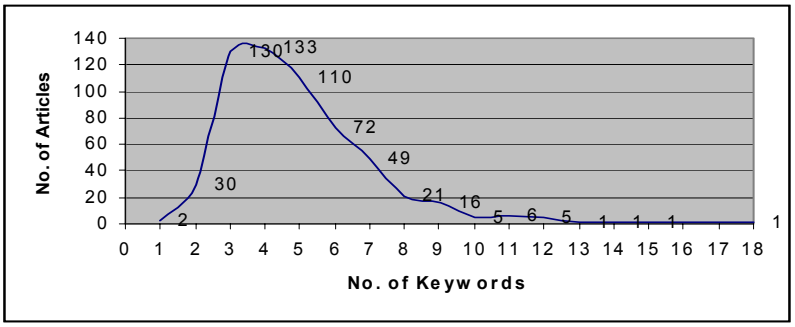

Figure 1. Keywords per Article Frequency Distribution

What Makes a Good or Bad Keyword?

During the process of researching, collecting, and analyzing the data for this study the authors spent large amounts of time "touching" the data. Entering in those 1791 keywords provided time for reflection on what makes a good or bad keyword. Based on this data set we find that poor keyword choices are based on ambiguity, or the improper use of "and", slashes "/", parentheses "()", and plurals.

Ambiguity

One of the problems with keywords is ambiguity (Gorla & Walker 1998). Ambiguity can best be described through an example. In our data set if one queries for articles that contain the keywords "System Design" they would retrieve three articles. Were there really only three articles published in *MIS Quarterly* about System Design? A closer inspection of the data shows that is far from the truth. Table 4 shows several variations of a System Design query and the number of articles with that keyword phrase.

Table 4. Keyword Query of "System Design"

Keyword Phrase	Number of Articles
System Design (original query)	3
Systems Design	10
System Designs	0
Systems Designs	1
System Analysis and Design	0
Systems Analysis and Design	4
Information System Design	2
Information Systems Design	8
MIS Systems Design	1
Participative System Design	1
System Design Methods	1
Expert Systems Design	1
Impact and Socio-Technical Systems Design	1
TOTAL	**34**

This table shows over a 10-fold increase in the possible number of articles about System Design. While it may be argued that some of these are more restrictive versions of system design (i.e. Expert Systems Design or Participative System Design), a strong case can be made that there is no difference between:

- System Design
- Systems Design
- System Designs
- Systems Designs

Yet depending on how you query you may get zero (system designs) or 10 (systems design).

The Inappropriate Use of "and", Slashes, Plurals, and Parentheses

While there are times when using "and" in a keyword phase is appropriate, such as "Systems Analysis and Design" where the phase is considered one topic; there are other examples when and separates two distinct ideas. In that case the words (phrases) really ought to be two separate keyword/keyword phrases.

Beyond the inappropriate use of the word and, worse yet, is the slash "/". The following keyword phrase "Technology and Organizational Behavioral/Political Issues" really becomes muddled in the eye of the knowledge seeker. A better solution would be to separate these two distinct phrases into two keyword phrases: "Technology and Organizational Behavior" and "Political Issues"

Plurals and parentheses may also impair the search capabilities of a knowledge worker. As an additional example, in our data set we found articles that had keyword phrases of:

- Decision Support System
- Decision Support Systems

- DSS
- Decision Support Systems (DSS)
- Decision Support

While there might be some support found for the argument that Decision Support varies significantly from Decision Support Systems, we argue that there is no difference in Decision Support System, Decision Support Systems, Decision Support Systems (DSS), and DSS.

Good Keywords

Given all of these examples of bad keywords, we suggest that a good keyword is:

- Concise (a keyword phrase that is 7-10 words long is not a good keyword)
- Descriptive (non-ambiguous)
- Popular (it has been used by several articles... Don't reinvent the wheel)
- Does not utilize hyphens, plurals, slashes, or parentheses inappropriately

Following the above guidelines, and those provided by your targeted publication, should ensure an article will be appropriately hit, using a keyword search mechanism.

Is There a Link Between Classifications and Keywords?

In addition to author provided keywords, most journals also specify categories for their articles. Rivard et al. (1988, 1993) have documented a classification scheme for IS research literature. Their classification scheme is categorized through an alphanumeric numbering scheme. This classification scheme documents some 1300 keywords produced by five leading IS journals over a period of five years (1987-1992). Their updated keyword classification scheme recognizes nine top level categories and 56 second level categories.

The Rivard et al. (1993) classification scheme forms the basis for the ISRL categories now used by *MIS Quarterly*. Prior to the development of this specific keyword classification scheme, *MIS Quarterly* relied on the ACM Computing Classification System. Before the adoption of the ISRL categories, *MISQ* employed two other classification schemes. The first, what we refer to as ACM64, is formally known as Categories of the Computing Sciences Classification System for Computing Reviews (1964 version). Prior to 1982, this classification scheme involved content indicators of two types (a) appropriate key words and phrases and (b) category numbers from a three-level tree.

In 1982 ACM published an entirely new classification system. That was the second system used by *MIS Quarterly*, and it continued to be used until the adoption of the ISRL classification scheme in June 1993; simultaneous with the publication of Rivard et al.'s (1993) updated article. For historical purposes the ACM classification schemes (last updated 1998) can be found at **http://www.acm.org/class/**, while the ISRL classification scheme can be found in Rivard et al. (1993) or electronically at **http://www.misq.org/roadmap/code/level1.html**.

All told, the data set for this article employs 613 unique categorization classifications, occurring 2287 times over 587 articles (21 articles did not provide classification categories). Articles that did not provide any classification categories were typically of type: issues & opinions, research notes, or introductions to special issues.

Like keywords, classification categories are meant to create similar groupings – in our context that would equate to grouping of similar articles. Upon a close inspection of the classification categories used on the *MIS Quarterly* articles we make two observations: 1) a very small set are used over and over again to classify the article in the most generic sense, and 2) like keywords, the bulk of the classifications are used a relatively small numbers of times. Table 5 shows the frequency of classification category usage.

Table 5. Frequency of Classification Category Usage

Number of times a category was used		Percent of Total
1	275	44.87%
2	111	18.10%
>2	227	37.03%
Total	613	100.00%

The most commonly used classification categories are shown in Appendix B. As with the top keywords, only categories that had 10 or more occurrences have been provided. For the reader's convenience, Appendix B also shows the classification code text description and the classification scheme that the code came from.

Table 6 presents some basic descriptive statistics on the classification category data, while Figure 2 charts the frequency distribution of categories per article.

Table 3. Descriptive Statistics of Categories per Article

3.90	mean
4	median
3	mode
2.08	stdev

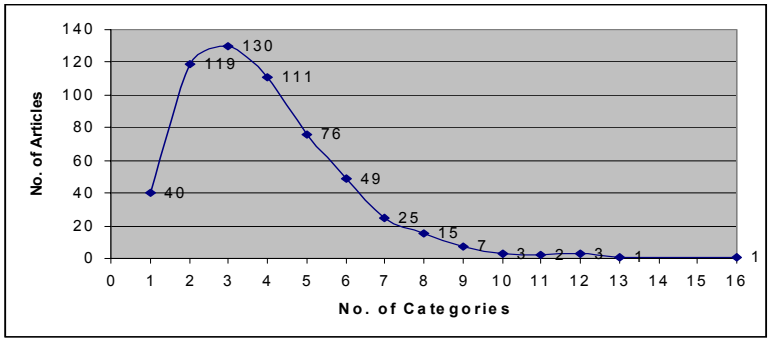

Figure 2. Categories per Article Frequency Distribution

Discussion

Given all this discourse about the variation in keyword phrases, we turn to the journal's policy on keywords. *MIS Quarterly* provides the following guidelines to authors concerning keywords:

> **Keywords**: Select 5-10 words or phrases to be used for indexing, in consultation with a technical thesaurus, if helpful. These might include important terms from the title, their synonyms, or related words. Do not use prepositions. Do not use hyphens unless the hyphenated parts are always treated as a unit. Use terms that are as specific as possible and whose meaning is generally known. For example, for an article title, "A Contingency View of Data Processing organizations," keywords might include: MIS management, managerial style, contingency theory;

Furthermore, *MIS Quarterly* also requires authors to submit appropriate ISRL categories for their article. Below are their guidelines to authors for classification:

> **ISRL** categories: category numbers from the ***Barki et al. classification scheme*** should be chosen as appropriate to the article. The scheme is also published in the June 1993 issue of ***MIS Quarterly.***

Clearly, these guidelines are fairly liberal, lending themselves to the propagation of the keyword limitations explored above. The recommendations not to use prepositions or hyphens help in selection of keywords; however, the policy could be stricter to include

parentheses and slashes as discussed above. Concerning the recommendation of 5-10 keywords, from our findings we see that over the 27 years of *MIS Quarterly*, authors have not heeded this policy. Assuming this policy was put in place recently, we re-examined the data set for the past four years (all 2000-2003 issues). Of the 64 articles published, 44 (or 69%) of them had 5-10 keywords.

If Not Keywords, Then What Other Techniques Can We Use for Extraction?

There is no doubt that keywords have helped reduce the time the knowledge worker spends attempting to retrieve knowledge. Those of us in academics would be lost without keyword search capabilities. However, is keyword searching as good as it gets, or can other tools and technologies be used to retrieve knowledge?

Various advances in the area of searching have helped information retrieval. For example, full-text searches now allow knowledge workers to search abstracts or entire bodies of text rather than limiting their querying capabilities to just the keywords provided by the author(s). Furthermore, advances in search algorithms including the integration of artificial intelligence (AI) and fuzzy logic can help. As an example Top-k selection queries can be used to find values without requiring exact matches (Chaudhuri & Gravano, 1999; Chen & Ling, 2002). While we recognize these advances, which largely are being developed in the computer science and engineering discipline, we suggest that much of the business and academic world have yet to incorporate these advances into their knowledge management systems.

We suggestion the use of "Knowledge Hierarchies," or classification schemes, as an alternative mechanism for information retrieval. Classifications schemes help knowledge workers in performing their search and retrieval tasks by putting keywords into a context. The context most often provided by category classification schemes is that of a tree-like hierarchy. While searching through the hierarchy individuals learn about the subject. Specifically, they gain additional knowledge of what is "near" their original term as they navigate the hierarchy.

Another benefit of having the structure of a hierarchy is that, if implemented electronically with modern technology, it allows for an entry point at the top, bottom, or anywhere within the body of the hierarchy.

As an example, suppose a doctoral student is doing research on the Technology Acceptance Model, or TAM. We already know what's going to happen if that student just does a keyword search. In our data set "Technology Acceptance Model" yields seven articles and "TAM" results in one hit. But was the student clever enough to search for "User Acceptance" which yields seven additional articles, including perhaps most importantly Davis' (1989) seminal paper that first proposes the technology acceptance model. Alternatively the doctoral student may have chosen to navigate the ISRL tree following "G" IS Usage, and "GB" Users, and there learned that GB02 articles discussed "User Attitudes" and GB03 discussed "User Behavior", both fundamentally associated with the Technology Acceptance Model. Querying GB02 finds an additional 19 articles that are TAM related for the researcher. Figure 3 depicts a portion of the ISRL Classification Scheme as a knowledge hierarchy.

Knowledge hierarchies have been prevalent throughout history. Information contained within file cabinets is typically sorted via folders that follow some sort of classification scheme. Furthermore, consider the structure of information stored on a computer hard drive. Whether using a Personal Computer employing Microsoft Windows or an Apple Macintosh, files are classified in folders structured by the end user.

A Suggested KMS Infrastructure for Information Retrieval

Relational database systems and full-text indexing technologies have served us well to date but additional technologies have been developed since and may aid us in fully realizing optimal information retrieval from our knowledge management systems. In this paper we propose an infrastructure based on dimensional modeling and multidimensional database technologies. By building knowledge management systems with multidimensional technologies, you gain the aggregation capabilities that are inherently hierarchical. Utilizing hierarchical tree structures allows for the same knowledge acquisition by navigation that you get by traversing classification schemes such as the ISRL, ACM, or SIPC codes. The beauty of these electronic hierarchies made available with multidimensional tools is that you can enter anywhere within the tree. There is no need for beginning at the root or leaf node, you simply select your entry point word phrase and navigate from there, drilling down for a more specific word phrase, or aggregating up to a more general category.

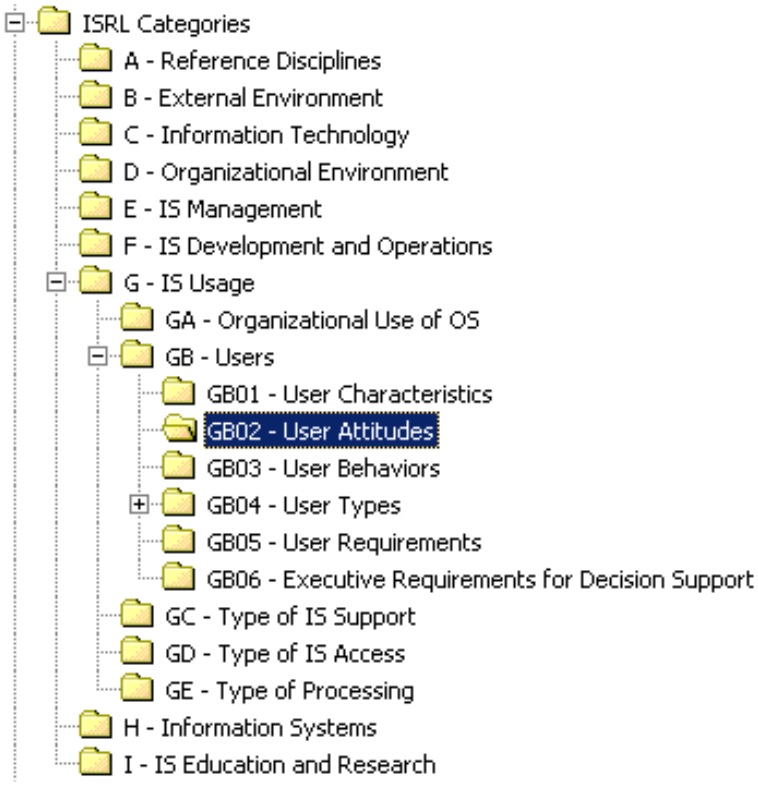

Figure 3. A Sample Knowledge Hierarchy (The ISRL Categories)

Furthermore, by employing these multidimensional technologies within the context of the knowledge management systems retrieval space, customizable "data marts for knowledge management" can be develop and utilized based on the dimensions that the knowledge worker cares about. As an example to carry this theme of academic journal management one step further, you can develop dimensions based on traditional topics such as author, keyword, journal, time or categories. Perhaps more interesting, however, is the ability to add dimensions such as reference disciplines, methods, variables, etc. To help visualize the power of this model consider being able to retrieve from a knowledge management system the answer to the following question, "What are all the variables that have been tested in TAM that turned out to be statistically significant?" Keyword searches are not capable of answering this question, but it is feasible to answer this question through the use of a dimensionally modeled KMS – a specialized information systems research literature dimensional data mart for knowledge management.

Testing a Knowledge Hierarchy Multidimensional KMS Model

An experiment is currently under way in which two knowledge management systems are being develop utilizing the same *MIS Quarterly* article data set. One KMS employs a relational infrastructure while the other employs a multidimensional structure. Subjects will then be tested on the accuracy of their information retrieval from the two systems. Statistical analysis will be employed on the data collected to determine if there is a significant difference in accuracy between the systems and to determine if a multidimensional approach utilizing knowledge hierarchies is superior for information retrieval.

Conclusions

In summary, we suggest that keyword searches are powerful and have helped the knowledge worker reduce the time spent seeking information, but they are not without their limitations, including: ambiguity, and poor choice of keyword phases, such as inappropriately using "and", slashes, plurals, and parentheses. Knowledge Hierarchies (or classifications schemes) are suggested as being more helpful, in that they force a selection of a keyword, thereby eliminating much of the ambiguity and redundancy.

Furthermore, knowledge hierarchies provide a navigable tree-like structure that enhances learning-by-viewing. It is with this concept of navigable hierarchies in mind that we propose a dimensional modeling approach and a multidimensional database infrastructure to help improve information retrieval success.

While this paper focuses primarily on academic inquiries to journal publications, this research has implications beyond the academic environment. Other contexts where this may be appropriate are tapping the expertise within the corporation and managing lessons learned or best practices. Additionally, the infrastructure suggested here could easily be adapted to accommodate the special needs of government management of resources.

Limitations

This study is not without its limitations. The data collected is only from one leading IS journal, and may not be representative of the entire IS field. While the authors recognize this limitation, we suggest that the data set is unique and valuable in the fact that it portrays the entire history of a journal that has paralleled the development of MIS as a disciple for nearly 30 years. Other studies may have included multiple journals, but they lacked the longevity of this study – that is other research on journals typically had an examination period of 5-10 years. Furthermore, the size of this data set – 608 articles, 1791 keywords occurring 2885 times, and 613 categories occurring 2287 times – provides a large enough sample for statistical analysis. Further analysis, as suggested by one review, could enhance the findings if the data was analyzed longitudinally for trends. As an example, further insight might be gained by breaking the data set into decades to see if the use of keywords changes and if so does it help the knowledge worker performing information retrieval tasks. This was addressed in a limited fashion, when we reanalyzed the data for the number of keywords used during the 2000s and found the trend was moving towards the recommended 5-10 keywords. Clearly there is room for more longitudinal analysis.

Future Directions

Ideally we would like to extend this data set to include additional journals (potentials include: Information Systems Research, Journal of Management Information Systems, Communications of the ACM, Information & Management, and others). We expect that the addition of these journals will give us further confirmation of the keyword patterns and limitations we found with the *MIS Quarterly* articles. Additional data in the form of various other articles will also allow us to reconfirm or suggest further modifications to the Barki et al.'s (1993) ISRL classification scheme. And finally, additional data might allow us to re-evaluate IS research domains and relationships as was demonstrated in the Swanson & Ramiller (1993) article.

Future research by the authors also includes the development of an experiment to test the accuracy of information retrieval from a traditional, relational knowledge management systems and one developed using a dimensional approach. Finally, contexts beyond the IS research literature are being explored, including case studies in the high-technology field and the military arm of the US government with respect to information retrieval from their 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.

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.

Chaudhuri, S. and Gravano, L. "Evaluating Top-k selection Queries," *Proceedings of the 25th VLDB Conference,* Edinburgh, Scotland, 1999, pp. 399-410.

Chen, C.M. and Ling, Y. "A Sampling-based Estimator for Top-k Selection Queries," *Proceedings of the 18th International Conference on Data Engineering* IEEE, 2002.

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.
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.
MIS Quarterly Website, (**www.misq.org**) last accessed 23 May 2003.
Neufeld, D.J. and Staples, S. "An Investigation of Information Systems Topic Coverage in IS versus Functional Area Business Journals," *Eighth Americas Conference on Information Systems*, Dallas, TX, Aug. 9-11, 2002. pp.1759-1760.
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.

Appendix A. Most Common Keywords

Keyword Phrase	Number of Articles Using Phrase
Decision Support Systems	33
information Systems	28
Management Information Systems	22
Implementation	19
IS Management	19
Systems Development	19
MIS Management	16
End-User Computing	14
Expert Systems	14
Group Decision Support Systems	13
Measurement	12
User Involvement	12
Group Support Systems	11
Management	11
Strategic Planning	11
Decision Support	10
Information Technology	10
Management of Information Systems	10
Planning	10
strategic Information Systems	10
Systems Design	10

Appendix B. Most Commonly Used Categories

Category	Count of Articles Using that Category	Description	Classification Scheme
K.6.1	90	Computing Milieux - MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS - Project and People Management	ACM'82-98
K.6	80	Computing Milieux - MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS	ACM'82-98
3.5	68	Applications - Management Data Processing	ACM'64
H.4.2	51	Information Systems - INFORMATION SYSTEMS APPLICATIONS - Types of Systems	ACM'82-98
J.1	47	Computer Applications - ADMINISTRATIVE DATA PROCESSING	ACM'82-98
K.6.4	43	Computing Milieux - MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS - System Management	ACM'82-98
K.4.3	41	Computing Milieux - COMPUTERS AND SOCIETY - Organizational Impacts	ACM'82-98
2.4	40	Computing Milieu - Administration of Computing Centers	ACM'64
H.4.0	38	Information Systems - INFORMATION SYSTEMS APPLICATIONS - General	ACM'82-98
K.6.3	33	Computing Milieux - MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS - Software Management	ACM'82-98
H.1.2	33	Information Systems - MODELS AND PRINCIPLES - User/Machine Systems	ACM'82-98
H.0	27	Information Systems - GENERAL	ACM'82-98
D.2.9	24	Software - SOFTWARE ENGINEERING - Management	ACM'82-98
K.6.2	22	Computing Milieux - MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS - Installation Management	ACM'82-98
GB02	19	IS Usage - USERS - USER ATTITUDES	ISRL
H.1	18	Information Systems - MODELS AND PRINCIPLES	ACM'82-98
D.2.1	17	Software - SOFTWARE ENGINEERING - Requirements/Specifications	ACM'82-98
H.1.1	16	Information Systems - MODELS AND PRINCIPLES - Systems and Information Theory	ACM'82-98
D.2.2	16	Software - SOFTWARE ENGINEERING - Design Tools and Techniques	ACM'82-98
H.4.3	15	Information Systems - INFORMATION SYSTEMS APPLICATIONS - Communications Applications	ACM'82-98
AI0102	13	Reference Disciplines - RESEARCH - RESEARCH METHODOLOGY - Case study	ISRL
GB03	13	IS Usage - USERS - USER BEHAVIOR	ISRL
2.41	12	Computing Milieu - Administration of Computing Centers - Administrative Policies	ACM'64
J.0	12	Computer Applications - GENERAL	ACM'82-98
AI0108	12	Reference Disciplines - RESEARCH - RESEARCH METHODOLOGY - Field study	ISRL
HA0301	12	Information Systems - TYPES OF INFORMATION SYSTEMS - DSS - Group DSS	ISRL
EL05	11	IS Management - IS MANAGEMENT ISSUES - IS TECHNOLOGY TRANSFER	ISRL
AI0104	11	Reference Disciplines - RESEARCH - RESEARCH METHODOLOGY - Empirical research	ISRL
1.3	11	General Topics and Education - Introductory and Survey Articles	ACM'64
4	11	Programming	ACM'64
3.3	11	Applications - Social and Behavioral Sciences	ACM'64
FD	10	IS Development and Operations - IS IMPLEMENTATION	ISRL
H.4.m	10	Information Systems - INFORMATION SYSTEMS APPLICATIONS - Miscellaneous	ACM'82-98
K.7.0	10	Computing Milieux - THE COMPUTING PROFESSION - General	ACM'82-98
K.0	10	Computing Milieux - GENERAL	ACM'82-98
2.11	10	Computing Milieu - Philosophical and Social Implications - Economic and Sociological Effects	ACM'64
2	10	Computing Milieu	ACM'64
AI0105	10	Reference Disciplines - RESEARCH - RESEARCH METHODOLOGY - Experimental research	ISRL