

Utilizing Human Categorisation Ability for Knowledge Management

James Sinclair

September 2004

1 Introduction

In 1998 and the years following, the Google search engine took the world wide web by storm, providing high-quality search results that were clearly superior to other commercial search engines at the time (Brin and Page, 1998). The algorithm that provided these high-quality search results relied on one simple assumption: “A hyperlink from page *A* to page *B* is a recommendation of page *B* by the author of page *A*” (Henziger, 2001). Although this assumption is not necessarily true in every case, the quality of results produced has forged Google’s acclaim as a household name. To “Google it” has become synonymous with searching the web.

Interestingly, the high quality search results were not the result of employing sophisticated data-mining techniques, nor did they depend on web authors entering quality metadata for their pages. The assumption was an observation of how humans interacted with an already existing system, namely the world-wide-web. In IT systems used in a knowledge management (KM) context, it is often not possible to make the same assumption regarding hyperlinks, since there may not be any hyperlinks used in the system at all. It is possible, however, to observe the way humans interact with the system.

The proposed research will entail observing human interactions with IT systems commonly used in KM contexts, such as Lessons Learned systems or Experience Management systems. This will be done with a view to improving the quality of searches made on these repositories. Of particular interest is how human beings categorise what they know in their own minds, and whether this can be utilized more effectively in the design of such systems. The aim is to improve the way information is organised, making it more accessible and promoting knowledge sharing.

Take, for example, a lessons learned system used by photocopier repair technicians. When a technician solves a particularly difficult issue with a photocopier, they make an entry into the system explaining the problem, its symptoms, and steps taken to solve it. This entry, along with others, is included in a report to be discussed with other technicians in their monthly meeting. The entry is also made available to product design engineers, so that design flaws are not perpetuated through subsequent photocopier models. When making a repair, the technician also has the ability to search through previously made entries to see if other technicians have encountered a similar problem.

When a technician makes an entry into this system, there is a high probability that they have some idea of how it relates to other entries that they have made previously. The aim of

the research is to determine whether this knowledge can be used to make entries in the lessons learned system more accessible and to facilitate sharing of knowledge through the system.

2 Literature

In the KM literature, information technology has a somewhat confused role (Barrett et al., 2004). Opportunistic software companies have “jumped on the KM bandwagon,” offering “KM solutions” and giving the impression that to implement knowledge management, all a manager need do is install their software. In response, knowledgeable writers in the field have sallied forth like preachers correcting a heresy, proclaiming that KM is more than software. They argue that “knowledge always involves a person who knows” (McDermott, 1999, pg. 105) and “Knowledge is not a thing that can be extracted from one cognitive system and fed into another” (Peschl, 2004, pg. 9). Hence, they argue, to talk about “knowledge capture” or “knowledge repositories” is misleading (Walsham, 2002). Knowledge, when it is “captured” becomes information, and so any “knowledge repository” is simply a mislabelled database. These writers often point to failed KM ventures based on large knowledge repositories that did nothing more than create “information junkyards” (McDermott, 1999) or intranets described as “large warehouses that nobody visits” (Walsham, 2001).

In light of these assertions, it would be easy to assume that information technology has no place in KM. These writers themselves point out however, that the problem is not so much with the technology itself, but with the human processes involved (Walsham, 2001; Barrett et al., 2004). These “human factors” are the subject of technology diffusion and technology adoption literature. Such human factors have been investigated by Pantano et al. (2002) and are the subject of another PhD thesis (Pantano, 2004).

It is not so much that “knowledge repositories” are useless (however incorrectly labelled), so much as that they have been implemented with a lack of understanding as to the human factors involved. There is also a lack of understanding as to how these tools are to be used and evaluated in supporting the KM goals of the organisation. In commenting on the KM literature Seaman et al. (2003, pg. 840) point out:

In general, this literature does not directly address the underlying technical infrastructure supporting the KM program, but instead mentions the use of already existing, off-the-shelf technologies and tools. Further, this work is concerned with the evaluation of a KM program once it is fully implemented and operational. It does not provide a way to evaluate pieces of the infrastructure for their suitability to support the KM goals of the organization.

It seems that those writing about information technology in the KM literature speak in very broad terms about inappropriate applications of technology. They argue quite persuasively that a different framework for thinking about how technology is used is needed to support KM, but show little imagination as to how the technology itself can be changed to support this. The research proposed here aims to address part of this problem by examining ways that the ability of human beings to categorise knowledge can be utilised more effectively in IT systems supporting KM efforts.

The problem of categorising knowledge appears with surprising regularity in the literature, yet it appears that little research is being done to address it. Linde describes the categorisation problem in this way:

In order to make [a knowledge artifact] useful, I have to do the work of knowing about the database, accessing it, and making the connection between my situation and the situation being narrated, using a category system provided by the designer which may have no relation to my way of categorising the problem. (Anyone who has attempted to use a manual to diagnose and fix a technical problem will be aware of this problem of categorisation of problems; all serious technical writers break their hearts over it). (Linde, 2001, pg. 165)

The problem appears again in McDermott (1999), where the author gives recommendations on how to best leverage knowledge. The following is from a sub-section entitled *Use the Community's Terms for Organizing Knowledge*:

Organize information naturally. [...] A good taxonomy should be intuitive for those who use it. To be “intuitive” it needs to tell the story of the key distinctions of the field. [...] Of course, this means that if you have multiple communities in an organization, they are likely to have different taxonomies, not only in the key categories through which information is organized, but also in the way that information is presented. [...] *The key to making information easy to find is to organize it according to a scheme that tells a story about the discipline in the language of the discipline.* (McDermott, 1999, pg. 114, original emphasis)

Although McDermott highlights the categorisation issue, he offers little advice as to how to go about developing an effective, intuitive taxonomy. Nor does he explain how to design the taxonomy in such a way that it is flexible enough to adapt as communities of practice change and the community's understanding of its own activities evolves. Linde (2001) also fails to provide suggestions as to how this issue of categorisation might be addressed, other than avoiding the use of “knowledge repositories” completely.

Lueg (2001) does address the issue to some degree in an article describing problems common to information management and KM. Lueg suggests the use of collaborative filtering, and social navigation. *Collaborative filtering* relies on “crowds of networked minds” providing “information concerning their likes and dislikes as ratings. These ratings are aggregated and are then used to compile recommendations for particular items” such as books or CDs. This approach has been implemented quite successfully on web sites such as the Amazon.com online store (amazon.com, 2004). *Social* navigation uses past usage history to assist users in navigating through virtual environments such as an intranet or database. A common analogy is that of following “cow paths” created by many people walking a similar route between buildings.

The collaborative filtering and social navigation approaches work well with systems that have large numbers of users (*cf.* “crowds of networked minds”). They also take advantage of the fact that in general people seem quite happy to enter information about their likes and dislikes of CDs, books and movies, etc. In many industrial situations, however, time is pressured and a system may not have the crowds of users needed to ensure the effectiveness of these techniques. Their focus is on usage after data has been entered into the system, rather than ensuring that data is effectively categorised at the point of entry.

3 Method and Plan

The first stage of the research will attempt to test two hypotheses:

1. Users entering a knowledge artifact into a repository-type system know more about the issue being described than is recorded by the system.
2. Some of this knowledge is indicative of how the issue being described relates to other entries in the system, that is, the knowledge is useful for categorising the artifact.

The aim of testing these hypotheses is to gain an insight into how users relate problem cases in their own experience to knowledge artifacts in a repository-type system. That is, to understand how knowledge in the participant's mind relates to whatever is stored in the computer. This insight will then form the basis for further investigations into how these relationships in the mind of the user can be utilised to enhance the categorisation of data in a repository. By enhancing the categorization of data, it is hoped that accessing it will become more intuitive, thus making the system more accessible. This is with the over-arching aim of facilitating and encouraging the sharing of knowledge within an organisation, supported by information technology.

Since the research is investigating human behaviour and understanding, testing the hypotheses is most appropriately done using qualitative research methods. Testing these hypotheses requires an in-depth understanding of individuals' cognitive processes in potentially complex technical domains. As such, surveys and statistical analysis are "less effective for generating understanding of the phenomena being researched" and "the results [of surveys] can be fairly superficial" (Kent, 2001, pg. 10). Conversely "It could be argued that human behavior is one of the few phenomena that is complex enough to require qualitative methods to study it" (Seaman, 1999, pg. 557). The use of qualitative methods is documented in both the KM and software engineering literature (Patriotta, 2004; Risku, 2004; Seaman, 1999; Seaman et al., 2003).

The hypotheses will be tested through a case study in an industry setting. It is proposed that the SIMPRESS system used at the Ford of Australia Metal Stamping Plant form the basis for initial investigations. SIMPRESS is a "knowledge capture" system initially designed to support die-building operations in the Geelong Metal Stamping Plant. There are a number of reasons for choosing this system:

- Firstly, the ANU Department of Engineering has a long-standing relationship with Ford of Australia, having undertaken several collaborative research projects over the last ten years. SIMPRESS was developed by the ANU Department of Engineering on behalf of Ford of Australia as a result of these collaborations. This should make gaining access comparatively easy due to the existing relations.
- Secondly, the researcher himself is familiar with both SIMPRESS, and the Stamping Operations Plant in Geelong. This means that less time will be required for the researcher to gain an understanding of the technical operation of the system and the processes surrounding it.
- Finally, SIMPRESS is an innovative system in that it processes more than just textual information. SIMPRESS allows users to upload images such as CAD drawings or digital photographs of a part and mark-up these images on screen. Thus, a user is able to take a digital photograph of a part, upload it, and draw a red circle (for instance) around the problem area, rather than attempting to describe the location of the issue in words. The presence of non-textual data in the system adds a dimension to the study that will

serve to enhance the results. Since human beings do not solely conceptualise knowledge in words, neither should an investigation into categorisation of knowledge be limited to textual data.

The research will be conducted through an iterative process of document analysis and semi-structured interviews. The entries already existing in SIMPRESS will be analysed, examining what *is* captured by the system, and thus informing interview structure. The users of the system will then be interviewed to gain an understanding of what is *not* captured by the system. This will involve both those who enter data into the system, and those who access the data after it has been entered.

Semi-structured interviews are chosen as a way of accessing narrative concerning the issues described in SIMPRESS. Narrative is discussed by Linde (2001) and Patriotta (2004) as an effective means to access tacit knowledge in an organisation. It is anticipated that the interviews will take place in the participants' places of work, and will be relatively short in duration (approximately five to ten minutes). The physical environment of the plant will make recording of interviews impractical, so detailed field notes will be taken after each interview.

Once collected, the data will be analysed to test each hypothesis. The first hypothesis may be tested by comparing entries in SIMPRESS with collected interview data. If the hypothesis is true, then the interview data will reveal information about cases not present in SIMPRESS.

Testing the second hypothesis will require analysing any interview data which reveals information not present in SIMPRESS. An analysis methodology similar to that employed by cognitive anthropologists will be utilised to investigate classification schemes and cognitive structures used by participants in relation to their expert knowledge (see Rossman and Rallis (2003) for a discussion of this approach).

3.1 Subsequent Research

As described above, the first stage of the project will involve a qualitative study into classification schemes used by employees with a high level of technical expertise in an industrial setting. Research subsequent to this study may follow one of two directions, depending upon the outcomes of the research. If the results suggest immediately applicable improvements to the design of repository-type systems, then the next step will be to develop a prototype system to evaluate their usefulness. On the other hand, it may be deemed necessary to undertake another case-study to verify the findings of the initial study.

Figure 1 shows a tentative outline of the various phases in the research. It is envisaged that a report of the findings will be written in a format suitable for a conference paper or journal article.

References

- amazon.com: 2004, Amazon.com online store. <http://amazon.com>, Viewed 13 September, 2004.
- Barrett, M., Cappleman, S., Shoib, G. and Walsham, G.: 2004, Learning in knowledge communities: Managing technology and context, *European Management Journal* **22**(1), 1–11.
- Brin, S. and Page, L.: 1998, The anatomy of a large-scale hypertextual web search engine, Vol. 30, pp. 107–117.

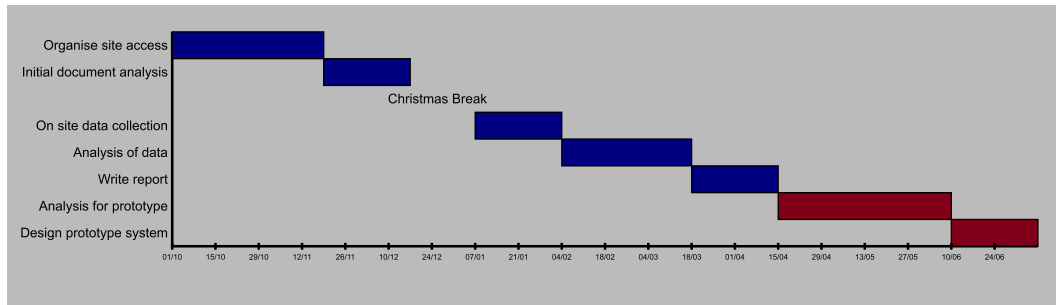


Figure 1: Overview of initial research

- Henziger, M. R.: 2001, Hyperlink analysis for the web, *IEEE Internet Computing* **5**(1), 45–50.
- Kent, R.: 2001, *Data construction and data analysis for survey research*, Palgrave, Houndmills, Basingstoke, Hampshire; New York.
- Linde, C.: 2001, Narrative and social tacit knowledge, *Journal of Knowledge Management* **5**(2), 160–170.
- Lueg, C.: 2001, Information, knowledge, and networked minds, *Journal of Knowledge Management* **5**(2), 151–159.
- McDermott, R.: 1999, Why information technology inspired but cannot deliver knowledge management, *California Management Review* **41**(4), 103–117.
- Pantano, V.: 2004, *An Interdisciplinary Model of International Technology Transfer*, PhD thesis, Australian National University.
- Pantano, V., Cardew-Hall, M. J. and Hodgson, P. D.: 2002, Technology diffusion and organisational culture: Preliminary findings from the implementation of a knowledge management system, *Proceedings of the IEEE International Engineering Management Conference*, Cambridge, UK, pp. 1–6.
- Patriotta, G.: 2004, On studying organizational knowledge, *Knowledge Management Research and Practice* **2**(1), 3–12.
- Peschl, M. F.: 2004, Structures and diversity in everyday knowledge: From reality to cognition, knowledge and back, *Organising Knowledge: Methods and Case Studies*, Palgrave Macmillan, Houndmills, Basingstoke, Hampshire; New York, chapter 1, pp. 3–27.
- Risku, H.: 2004, Situatedness and artefacts in expert knowledge and activity, *Organising Knowledge: Methods and Case Studies*, Palgrave Macmillan, Houndmills, Basingstoke, Hampshire; New York, chapter 2, pp. 28–42.
- Rossmann, G. B. and Rallis, S. F.: 2003, *Learning in the Field: An Introduction to Qualitative Research*, Sage Publications, Thousand Oaks, California.
- Seaman, C. B.: 1999, Qualitative methods in empirical studies of software engineering, *IEEE Transactions on Software Engineering* **25**(4), 557–572.

- Seaman, C. B., Mendonça, M. G., Basili, V. R. and Kim, Y.-M.: 2003, User interface evaluation and empirically-based evolution of a prototype experience management tool, *IEEE Transactions on Software Engineering* **29**(9), 838–850.
- Walsham, G.: 2001, Knowledge management: The benefits and limitations of computer systems, *European Management Journal* **19**(6), 599–608.
- Walsham, G.: 2002, What can knowledge management systems deliver?, *Management Communication Quarterly* **16**(2), 267–273.