# Capturing Tacit Knowledge using Recommender Systems

Dick Stenmark
it.dixi@memo.volvo.se

Volvo Information Technology
Department 9530, HD1N
SE-40508 Göteborg, SWEDEN

IT & Organisation
The Viktoria Institute
Göteborg University
P. O. Box 620
SE-40530 Göteborg, SWEDEN

## Abstract

*People are able to determine whether a given document is interesting or not just by glancing through it. However, when asked to make explicit the rules upon which such a decision is based, they are unable to do so. Based on empirical findings I argue that our interests are examples of tacit knowledge, and that agent-based retrieval systems may be used to capture and externalise these interests, thus making otherwise elusive knowledge tangible. By using agent-based retrieval systems, two otherwise troublesome obstacles are avoided; interests may be defined by examples, and an incentive for creating and maintaining a search profile naturally exists.*

**Keywords:** Tacit knowledge, knowledge management, agent-based retrieval systems, intranet.

**BRT Keywords:** AL, AA, GA

# Introduction

Epistemologically, knowledge may be split along several dimensions. One way is, as suggested by Polanyi (1966), to distinguish between tacit knowledge and explicit knowledge. Though others have since developed this separation further, (e.g. Blackler (1995) speaks of embodied, embedded, embrained, encultured, and encoded knowledge), I shall stay with Polanyi's definition for the scope of this paper. Thus with explicit knowledge, I will mean knowledge that has been captured and codified into manuals, procedures, and rules, and when using the phrase tacit knowledge, I refer to knowledge that cannot be easily articulated and thus only exists in peoples hands and minds.

An interesting but also troublesome property of tacit knowledge is the inherent tension between its value on the one hand and its elusiveness on the other hand. The high value stems from the fact that most of the body of knowledge is made up of things we know but are unable to express. Polanyi, who was one of the first to discuss tacit

knowledge, explains: “We can know more than we can tell” (Polanyi, [1966] 1998, p.136). Leonard and Sensiper (1998) go even further by stating that "we can often know more than we realize" (p.114). One of the objectives of Knowledge Management is to bridge the gap between tacit and explicit knowledge, and any technological solution that could assist in this process would thus be highly appreciated.

The elusiveness of tacit knowledge can be derived from at least two reasons; we are ourselves not fully aware of it, and there is a lack of incitement on the individual level to make it explicit. Firstly, tacit knowledge “incorporates so much accrued and embedded learning that its rules may be impossible to separate from how an individual acts” (Davenport and Prusak, 1997, p.70). In our daily activities, our tacit knowledge informs our activities without us thinking of it as knowledge. We know how to ride a bike, or what cinnamon smells like, but we cannot document it in a manual, nor explain it to others. Secondly, our knowledge is something that resides within us, and manifest itself though our actions, and we therefore do not need to document it for our own sake. We just use it. Should we have to express our tacit knowledge in words, it would not be for our own sake but for the benefit of someone else in our organisation or community. Grudin (1987) has argued convincingly that situations where one is forced to do the work and someone else gets the benefit very often result in failure.

Orr (1996) describes how an organisation's view of how work is carried out can contrast sharply to what it really takes to get a job done. Though we have our formal job descriptions, these are seldom enough to account for the actions we perform during a working day. Our interests as professionals often make us elaborate within, and sometimes even outside, our role definitions. Much of our daily office activities are thus governed by professional interests that dictate what reports we read, what documents we write, what discussions we engage in, and what we search for on the web.

I suggest that our interests are instances of tacit knowledge. Though we may be unable to produce an exhaustive definition of our interests, we usually have no problem in determining whether any given document is interesting or not. Therefore, we intuitively know what we are interested in when we see it but we are unable to make our interests explicit for others to learn. Suchman (1987) observes that tacit knowledge enables us to take actions that are situated in particular social and physical circumstances, and that tacit knowledge thus is contextual bounded. In an office setting as the one described in this paper, our interests, and the actions they give birth to, are limited to a professional context. If we could capture some of those activities and derive our underlying interests, we might be able to externalise part of our tacit knowledge, and thus make it -- if not explicit --at least “touchable”. Such a possibility would be useful to an organisation, as it would enable the sharing of this value resource.

I argue that agent-based web retrieval systems can be used to solve the two problems mentioned above; it helps articulate tacit knowledge, and; it creates an incentive to try to do so. By identifying certain documents as interesting, an agent-based retrieval system could maintain a dynamic profile that represents some of my tacit knowledge without requiring explicitly defined keywords or manually updated records. Since this profile is used to provide me with information that is more accurate and search results that are more precise, a natural incentive exists for me to give feedback and thus cultivate the profile. The resulting profile, I will try to show, represents part of my tacit knowledge, which in a sense becomes tangible.

Research concerning agent-based retrieval systems has focused mainly on user-to-object or user-to-information objectives, but has sometimes also addressed the user-to-user considerations. See Resnick and Varian (1997) or Fagrell and Ljungstrand (1998) for

references to various recommender systems and their implementations. No one, however, has approached agent-based retrieval systems from a knowledge management perspective; i.e. discussed what knowledge governs the individual activities and how tacit knowledge may be put to use in the community. My work contributes by proposing an interpretation that explains how tacit knowledge is activated, and how it may be made tangible. Since my work is focused on IT usage, I have studied people rather than the technology itself. To study whether agent-based retrieval systems could assist us in defining our true interests, I needed an intranet retrieval application prototype. Therefore, I used a commercially available tool to implement the prototype, but the choice of tool was not significant for the research - any equivalent product would have worked.

In the next section I will explain my research methodology and describe the domain in which I performed my empirical study. I will then describe the fundamental features of the prototype and continue with a report of the users' experiences and comments. Thereafter, I discuss these findings and suggest more general interpretations before concluding with a summary.

# Research settings and method

The empirical fieldwork took place at Volvo Information Technology, an IT service company within a large European corporate group. Their intranet consists of some 450 web servers and has approximately 400,000 documents. Most of the content is official or semi-official information, such as department presentations, project reports, frequently-asked-questions (FAQ's), and online help material.

I invited approximately 80 users to participate in the study, which ran from August to November 1998. Most of the users were Volvo IT employees but their backgrounds varied from administrators and content providers to system developers and technicians. All had thorough computer experience.

Forty-eight of those did actually register and participated in the test. All users were invited to a 2-hour introduction meeting, where I explained the purpose of the research, the design of the application and how to operate it, how to register and login, and how to set up and run individual agents. I also asked the participants to keep informal records of particular incidents that they considered worth noting, and informed them that I was going to contact them during or after the test to collect their viewpoints. The seven users that were unable to attend either of the three introduction meetings received the above information via email.

User experiences have been collected in several ways. All users were invited to a group interview but only eight showed up. The remaining users were then sent an email questionnaire, which again only some answered. After a first analysis, based on the so far received answers and the application log files, I conducted seven semi-structured qualitative interviews, each lasting between 28 and 66 minutes.

# Fundamental features of the prototype system

My research interest was not to develop and study agent-based retrieval systems per se, but rather to examine how such a technology could be used in a new and innovative way. To speed up the development process I therefore wanted to build on existing software tools, if such where available. Recommender systems are able to anticipate what items a

user is likely to be interested in and can thus, in a hopefully intelligent way, recommend such items. How this “anticipating intelligence” is implemented varies from product to product and is not relevant to this paper. Academic research as well as the success of commercial products has shown that such systems do work and we may safely assume this to be true also in this particular case.

## Autonomy's Agentware

After some consideration, I decided to base my prototype on Autonomy's Agentware products. The heart of the product is the Dynamic Reasoning Engine (DRE) which uses neural networks and advanced pattern matching techniques to identify text patterns in profiles and look for similar patterns in other profiles or web documents. Each web document is synthesised into a 0.5K digital representation, a "fingerprint", and the characteristics that give the text meaning are determined. Once the fingerprint signature has been created, the DRE can perform concept matching (e.g. finding documents relevant to each other), agent creation (e.g. setting up agents that can find relevant documents), and agent retraining (e.g. adapt the agent to a set of relevant documents). I will not describe Agentware in depth but refer the interested reader to the product white paper (Autonomy, 1998).

## Design decisions

My prototype was designed and implemented to support the following:

1. Offer individual agents that could be set to find intranet documents based on a user profile, e.g. a richer representation of an interest than merely a keyword-based query.
2. Provide mechanisms to enable refinement of the agents based on positive user feedback on retrieved documents.
3. Enable users to locate colleagues with similar assignments and organisational roles by matching user job descriptions and thus facilitating the creating of online communities.
4. Display users with similar interests by matching their agents.

Below, these features are described in more detail.

#### 1. Creating agents

Since each user was to be offered personalised agents, they had to identify themselves by logging in. Figure 1a below depicts the initial screen with the private login section to the right. Once inside, the users could create an agent, give it a name, and assign it a task. The task corresponded to a search engine query, but was not limited to simple keywords and Boolean connectives since many users found such syntax difficult. Instead, the task was to be expressed in natural language and the best results were achieved when the users cut and pasted (a large chunk of text from) a relevant document and asked the agents to find more similar documents.

After the agent was created, it was represented on the screen as an image. As seen in figure 1b, the users have four options for each agent; delete it, edit it, find similar agents, or check the result.

**2. Agent refinement**

The search results from the agents were displayed in a simple list, similar to those generated by most search engines, and by clicking on the associated hyperlinks the documents were retrieved. When the user had read and verified that one or more of the returned documents were indeed relevant, the user could provide the agent with positive feedback by marking the document(s) and clicking the retrain button. See figure 2b for an illustration. The digital signature of the agent was then merged with the signature(s) of the selected document(s) and the result became the new agent signature, replacing the previous one.

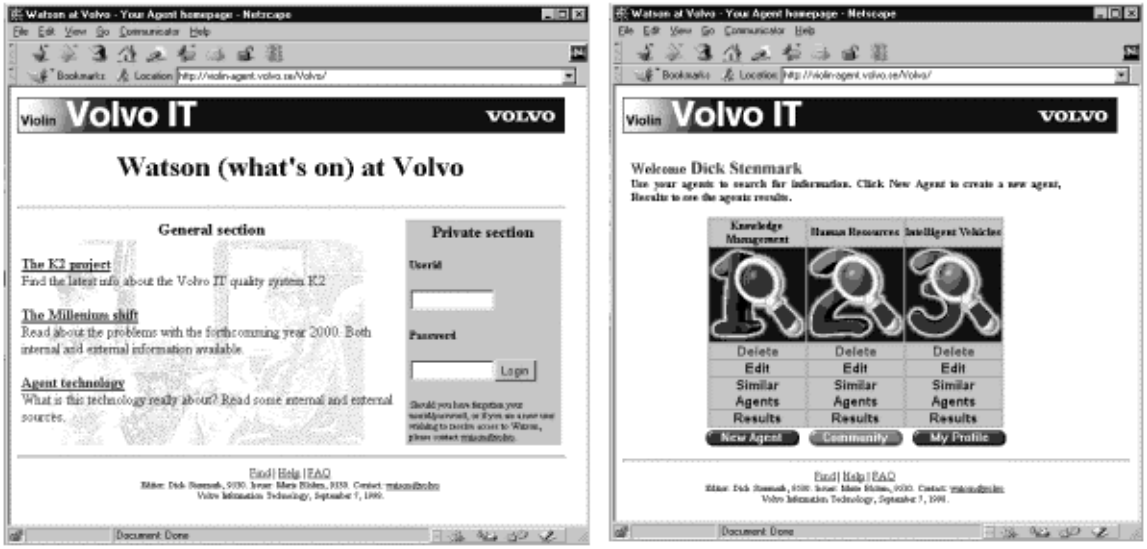


**Figure 1a and 1b. The login screen and the agent screen, respectively.**

**3. Profile-based communities**

New users were also suppose to create a user profiles in which they were to describe their professional interests or work responsibilities in a free text fashion. If a user already had a CV stored elsewhere, it could be copied into this field. The profile, once saved and stored, was then converted to a digital signature, and when clicking on the Community button, the user profile signature was matched with that of other users and the resulting users were listed on the screen as seen in figure 2a below. The user could now display the email address or the profile of any found user by clicking the corresponding hyperlink, and had the opportunity to contact him or her. My intention with this feature was to make the users aware of each other's presence and thus facilitate the emergence of online communities.

**4. Finding similar agents**

The Similar Agents feature was a rather late idea added more or less because it was relatively unproblematic to implement. I originally planned to let the users be able to search for and find similar agents to have them cloned by copying them to their own private area. This, I believed, would help new and inexperienced users to get their agents to a decent quality level quicker. However, this extra functionality was not implemented due to lack of resources and the only feature offered to the users during the test were the option to find other users with similar agents. The result of such a search was identical to the list produced by the Community feature.

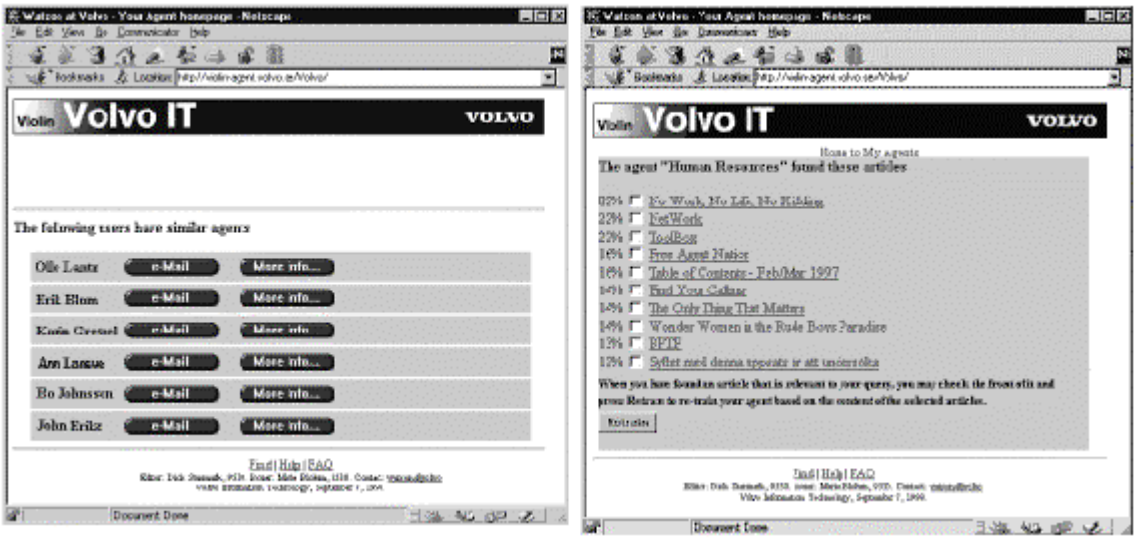


**Figure 2a and 2b: The list of users with similar agents and the list of returned relevant documents, respectively.**

# User experiences and comments

The users have typically used the application often, sometimes heavily, during a couple of days and then stayed away for some while before returning for the next session. Usage was especially high right after the release of the application, and then declined slightly before settling on a stabile level throughout the rest of the test.

## Mixed feelings and results

Overall, the user reactions were very positive. The respondents said they believed in this technology and considered it to be “an extremely important asset” with a “great potential”. The most frequently reported reasons for these beliefs were that it was “easier to construct queries” and that it “saved time not having to search”. One user put it this way; “In the future we're gonna be bombarded with even more info and this may be the only way to stay ahead”.

More specifically, several users expressed their appreciation of not having to come up with descriptive keywords, since “they do never fully contain the meaning you have in mind any way”, to use the words of one interviewee.

Despite the general claims that these sorts of Information Retrieval (IR) agents were welcomed and appreciated, the users had experienced mainly negative actual results. Many users reported what they referred to as “strange” or “unexpected” document matches. “[It is] hard to get something useful out of it. After retraining it with relevant documents it comes up with nothing” as one user put it. However, the users tended to blame these bad results on their own inability rather than on the application. One user having received very little useful information said, “The rather shallow results may depend on me not using the right words. Otherwise, I like the idea. Keep improving!”

### New practice requires training

The prototype differed from how conventional web IR tools, such as search engines, are operated, and the users had some problems adjusting. A common problem was that the users "overtrained" their agents, i.e. they kept retraining the agents until they became so specialised they did not match any document on the intranet. Some users actually commented that they felt that the way the agents worked had to be better understood by them in order to be able to interpret and anticipate the outcome of the search. The users who attended the introduction meetings appreciated the information given there and claimed that it helped them use the tool in a better way. The answers from the users not attending the initial meetings also clearly indicated that training was required. One non-attending user said: "It took some time before I understood [how to set up the agents] … I believe the instructions should be even more clear, so that one understands what the agent is and how it works".

### Little interest for static job descriptions

Not many of the test users exploited the Community feature, i.e. the feature that locates user with similar work-description profiles. The reasons given for not using the function were that the users already knew enough people doing similar jobs or that most users with similar profiles worked at the same department as the respondents. The respondents were not too interested in finding like-minded colleagues. As one user put it "What's the use of hooking up with people doing the same stuff I do? […] It would probably be better to team up with those who knows stuff I don't know." This last remark was an opinion shared by several users, who suggested that an opportunity to search for users with complementary profiles would have been more useful. Of those who actually did try the Community feature, all but one considered it to be working. One user, however, claimed to have been connected to people with whom the he had nothing in common. "This was clearly a bug" were his words.

### Action-informed similarity more appreciated

The Similar Agents feature was however much more frequently used, though it generated much the same sort of knowledge as did the Community feature. "It's really interesting to see who else is searching for this sort of thing", one of the users commented. Many respondents reported that they were surprised to find certain people sharing their interests, or that the Similar Agents feature returned users not expected to be interested in a particular topic. These comments were not uttered in a negative way. On the contrary, the users regarded these results as useful new insights and no one questioned the correctness of the results.

## Discussion

The users clearly preferred exemplifying their interests by pointing to relevant documents rather than having to invent clever keywords. This, I argue, is because the former involves tacit knowledge while the latter requires a translation to explicit knowledge. I further suggest that this distinction between tacit and explicit knowledge also explains the different ways in which the Community feature and the Similar Agents feature were used.

## Espoused theory in formal job descriptions

The Community feature, which is based on explicit knowledge, has not been used much at all. There are several possible reasons: Firstly, this research is carried out on a corporate intranet and not on the World Wide Web. The whole organisation may see itself as a community of which the users are already members, and the need to be associated with sub-communities may not be particularly strong. This I find not very likely. Secondly, conventional office tools, e.g. word processors or email programs, do not support a community concept and the users are thus not familiar with this way of working. They do not see themselves as community members, and do not appreciate the full potential of the concept. Tangible business benefits may first have to be experienced for this new way of working to be appreciated. This alternative seems more plausible than the first. However, none of these two is a strong candidate.

The third and most feasible explanation is that the Community feature was built on static profiles provided by the users themselves to mirror the official responsibilities placed upon them by the organisation. These profiles are presumably already known to the members and experienced as fictitious, since people are often viewed as performing their jobs according to their formal job descriptions though everyday practice provides evidence of the opposite, as shown by Brown (1998). This is consistent with the findings of Argyris and Schön (1974) who refer to the worldview and values that people believe their behaviour is based on as "espoused theory" as opposed to "theory-in-use". The users rightly or wrongly assume that they know what the Community feature will return and they dismiss it as uninteresting.

## Practice considered more relevant

The Similar Agents feature is different from the above in that it does not rely on static profiles but on the tacit knowledge of dynamically retrained agents, initially created with a totally different purpose than the static profiles. If the prompt "Enter your profile" connotes a question equivalent to "what is your official job description? " the agents are instead created for personal benefit only and no official considerations are taken into account. True and real interests govern the choice of topics, which makes these search profiles more "believable" than the previous job describing ones. The most notable observations from the interviews are that when matching job profiles built on explicit knowledge and espoused theory of work, the user being linked to unexpected colleagues referred to the result as "strange" in a negative meaning. At the same time, the users matching agents built on tacit knowledge and practice commented similar results as "interesting" in a positive meaning. The tacit theory-in-use is obviously regarded as more trustworthy.

## Comments on future research

In a future prototype, an added feature would be the possibility to explicitly search for a specific competence by entering a description and match it against both agents and user profiles. To be able to find this sort of knowledge is, however, only a first step; it only helps identifying people within the organisation - it does not prevent these people from leaving the organisation nor guaranteeing that they will have time to share their knowledge on request. Davenport and Prusak (1997, p.81) observe that "mapping who knows what in an organization creates an essential knowledge inventory, but does not

guarantee the ongoing availability of knowledge". To achieve such permanent knowledge capturing and storing, other measures that fall outside the scope of this paper must be deployed.

# Summary

I have argued that agent-based retrieval system technology could act as a facilitator in the knowledge managing process of capturing tacit knowledge on an intra-organisational web. The benefits of such an approach are mainly two: i) the otherwise hard to solve problem of being able to produce an exhaustive definition of one's interests is replaced with the much simpler task of determining whether a given document is interesting or not, and ii) since a good profile results in more accurate information, a natural incentive to maintain the profile by giving feedback exists.

Previous research on agent-based retrieval system has studied how to connect users with information or users with other users. While this study shows that both these goals may be achieved simultaneously, it also introduces a third, and until now unnoticed, aspect of agent-based retrieval systems. Profiles based on the tacit knowledge of practice are conveyed as more trustworthy than the espoused theory-based job description profiles. My contribution is the suggestion that profiles based on the tacit knowledge of our interests and identified by practice are conveyed as more trustworthy than the espoused theory-based job descriptions. The former profiles can be used to facilitate the externalisation of tacit knowledge in form of user interests to searchable knowledge in a low-intrusive way.

# Acknowledgements

Thanks are due to Volvo Information Technology for their support during this research, and to Autonomy Nordic for letting me use their software. I am also grateful to Jan Ljungberg and Magnus Bergquist at the Viktoria Institute for useful comments on earlier versions of this paper.

# References

Argyris, C., and Schön, D. (1974), Theory in practice: Increasing professional effectiveness, Jossey Bass, San Francisco, CA.

Autonomy, Inc. (1998), "Autonomy Agentware: Technology White Paper". Available at http://www.autonomy.com/tech/whitepaper.pdf [July 1999].

Blackler, F. (1995), "Knowledge, Knowledge Work and Organizations: An Overview and Interpretation", Organization Studies, vol. 16, no. 6, pp 1021-1046.

Brown, J. S. (1998), "Internet technology in support of the concept of "communities-of-practice": the case of Xerox", Accounting, Management and Information Technologies, no. 8, pp. 227-236.

Davenport, T. H. and Prusak, L. (1997), Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, Boston, MA.

Fagrell, H. and Ljungstrand, P. (1998), "Make an Agent and you shall find: An Intranet Recommender System", In N. J. Buch et al. (Eds.), Proceedings of IRIS21, Department of Computer Science, Aalborg University, Denmark.

Grudin, J. (1987), ”Social Evaluation of the User Interface: Who does the Work and Who gets the Benefit?”, In Bullinger, H.-J. and Shackel, B. (Eds.), Proceedings of INTERACT '87, Elsevier Science Publishers, North-Holland.
Leonard, D. and Sensiper, S. (1998), “The Role of Tacit Knowledge in Group Innovation”, California Management Review, vol. 40, no. 3, pp. 112-132.
Orr, J. (1996), Talking About Machines: An Ethnography of a Modern Job, Cornell University Press, USA.
Polanyi, M. (1966), The Tacit Dimension, Routledge and Kegan Paul, London.
Polanyi, M. (1998), “The Tacit Dimension”, In Prusak, L. (Ed.) Knowledge in Organization, Butterworth-Heinemann, Boston, MA.
Resnick, P. and Varian, H. R. (Eds.) (1997), “Recommender System”, In Communications of the ACM, vol. 40, no. 3, (entire issue).
Suchman, L. (1987), Plans and Situated Actions: The Problem of Human-Machine Communication, Cambridge University Press, Cambridge, UK.