Knowledge Management with Snapshots

Hilda Tellioğlu

Vienna University of Technology
Institute of Design and Assessment of Technology
Argentinierstrasse 8/187, 1040 Vienna, Austria
hilda.tellioglu@tuwien.ac.at

Abstract. This paper presents the concept of snapshots for knowledge management in distributed cooperative work environments. Snapshots are used to record important stages of collaborative project work, provide mechanisms for indexing, retrieval and visualization of relevant data for users in form of suggestions. They combine meta-level information with content and enable additional tagging and commenting. A prototype *SUGGEST!* has been developed to implement and evaluate the functionality and usability of such a knowledge management system.

Keywords: Knowledge management, information retrieval, snapshots.

1 Introduction

In a cooperative work environment, companies need to establish computer supported knowledge management. Project members have to find out easily who to contact, whom to ask certain questions and who to cooperate with in concrete projects. Experiences in previous projects need to be used to decide which company to collaborate with in future. Some of partner networks established are sustainable. Some are only temporal. In case of a real project setup, if one member of the network is not available, the most of the plans need to be modified. Sometimes a new partner must be found very fast.

Not only at the beginning of a project in globally networked settings, but also during ongoing projects, it is usually very difficult to capture, retrieve and manage information about people's knowhow, skills, availability and experiences. In most cases, the worth full information about past projects is not really existing or available. Several questions need to be answered: What did happen in a certain past project with a similar context? Who was involved in which part of a project? Who (internal or external) could cooperate without any problem? And so forth. Knowhow and experiences of past could help to avoid repeating mistakes done so far. Unfortunately there are still problems to support people in companies in organizing themselves and in creating their network due a specific project setting.

Although this problem is very well defined, there are no real solutions to it. Some attention has been given to technology development. Research in the area of semantic web has been done, e.g., about large-scale information extraction [1] [6] of entities

and relationships from semi structured and natural-language Web sources [10] or about extracting valuable data from hidden Web sources by using a combination heuristics and probing with domain instances as well as by using supervised machine learning technique [7]. The main idea was to use Web as a comprehensive knowledge base. Others used ontologies to manage multimedia knowledge [5]. Takahashi et al. [9] tried to support designers by providing a web-based knowledge database that selected the information mainly from news sites and reduced the useless information for design by applying a filter.

We found other research work focusing on concepts and theories. Speer et al. tried to acquire common sense knowledge by using a game based interface [8]. Chung and Hossain [2] developed a theoretical model based on social network theories and the social influence model to understand how knowledge professionals utilize technology for work and communication. They investigated task-level and communication-level ICT use.

Finally there is research on combining conceptual and functional level. Georgalas [3] advocated that knowledge repositories support strong mechanisms for the management of meta-level information and system specifications. They tried to integrate meta-level knowledge of systems and data at a conceptual level, before they integrate knowledge and software components at a functional level, analyzed in an environment called DELOS [4].

All these approaches are not really addressing the problem of knowledge management in organizations. It is still not possible to establish an easy-to-setup and easy-to-use system in organizations to capture, store, index, retrieve and visualize relevant project and product information in a lightweight. Only relevant data about and in projects need to be recorded and linked to each other. Additionally links can be tagged to provide another level of information about relations between people or artifacts or objects in the network of information captured.

In this paper the theory if knowledge management by means of snapshots is introduced to meet the requirements mentioned above. A snapshot freezes a moment in time, photographic or electronic. If a computer creates snapshots, you can time travel in your computer system. You can jump to a former state of the system. So, a snapshot can help capture properties and relations between components of a system. A snapshot can be used to establish knowledge management in an enterprise. We can record important stages of cooperative project work, provide mechanisms for indexing and tagging, in order to be able to retrieve them in the future. With a configurable visualization system users can make sense and use of this information.

This paper tries to design such a system, first theoretically and then prototypically. In an example, based on ethnographic research in automotive industries, it illustrates how to proceed in dealing with snapshots in a coordinative work environment. Requirements to a supporting system are described in a prototype by showing which information is needed to support creation, access and maintenance of a knowledge base for a work group.

In the next section we will describe the use case setting in which we gathered the requirements and data that we used in our prototype. In the section 3 we introduce the concept of snapshots for knowledge management, which follows the description of the prototype in section 4, before we conclude the paper.

2 Our Case

We carried out ethnographic studies in a European STREP project called MAPPER (Model-based Adaptive Product and Process Engineering) (IST-016527)[1] at three different industrial sites. One of them is KeyA that we refer in this paper to. KeyA is a company that produces car parts like gearshifts, head strains, and seat heating for automotive industry. It has several branches all over the world. The projects are multinational, involving different people from different branches depending, on the one hand, on people's skills and knowledge, and on the other hand, on the production facilities at the site. The geographically distributed way of project organization makes computer-supported communication and collaboration necessary. Meetings are arranged regularly to overcome the distance between distributed project members. Common information spaces have been established to enable central management of project documents. We could observe several meetings and carry out in-depth interviews with some of the key actors during four visits between 2005 and 2008.

Project members of all sites were used to participate regular and ad-hoc *meetings*. By means of teleconferencing facilities and sometimes screen sharing, single open issues were discussed by asking responsible persons the status of the work progress. In some meetings, suppliers or customers were present. The main artifacts used in all kinds of meetings were *to-do lists* that actually belonged only the project managers, who were the moderators of the meetings. Sometimes, *emails* were mentioned or quoted when there were misunderstandings, gaps of communication, uncertainty about the status of some issues. Emails provided different communication paths within and across distributed groups. They were used as information holders with attached documents like text files, presentation slides or spreadsheets related to the current project. They were also applied just for arranging, clarifying or negotiating open issues. Project members were actively using their emails, e.g., to search for information in their attachments, where the information in the email's body was used to identify the right document. Mailboxes were used to manage emails. No one would like to delete an email exchanged in a current project. Emails helped keeping track of data and document transfer between cooperating people. Documents created at KeyA were mainly *text files, spreadsheets* or *presentation slides*. Besides to-do lists, MS Power Point slides were the main project management instrument. For certain purposes project members were using other artifacts as well, like a groupware system to manage the emails and files on a common information space, a PLM system for managing product lifecycle, chat logs, product catalogues, product descriptions, test reports, etc. No configuration management environment has been established.

In our ethnographic study at KeyA[2] we could identify the need for organizational knowledge recovery what we illustrate here with two cases. Several questions about projects, persons and artifacts came up and could not be answered in a timely manner. In several meetings, there were questions about who to contact to carry out a certain activity or to prepare information to decide a certain step in the project, what the problem in a previous similar project was, who was involved in certain projects and problem cases, etc. In one case, it was not clear what really happened a couple of

[1] http://mapper.eu.org/

[2] The studies were carried out by Gianni Jacucci, Hilda Tellioğlu and Ina Wagner.

years ago in an ongoing project (P). They missed to send a document to the customer (C) and the same customer was now asking for it. Unfortunately, the project team has been changed, persons involved at that time were not present any more, no one had a clue who to ask or where to find the document the customer was waiting for. There were several meetings to plan a strategic solution to this problem and try to reduce the financial damage for the company. The second case is about missing information on part id's needed in IMDS system in use within the automotive industry in compliance with a EU directive, which requires companies to provide a detailed specification of the materials used in their parts. The following excerpt describes some of the difficulties involved in information retrieval with and around IMDS:

> A[3] checks her email – the first is a request from their office in Detroit concerning IMDS calculations – a colleague needs details for different heating wires. The email is about not finding the weight for all the wires and has an MS Excel sheet attached. A has a file on her drive – she talks about having to copy the files all the time and sending them back and forth, she opens a table, noting down numbers on a post-it. On her list there are two heating wires but colleague D looks for more – A sends him the ID for one of the wires and then tries to call the responsible for heating wires to ask her. A continues searching in another list and still cannot find the item. In the meantime A picks up another email, which is also about IMDS. A explains: "No one is able to give us the right ID for some parts – the problem is that we deliver one part directly to the car producer and one to the seat producer and each has their own part ID – each supplier to LL has different numbers and at LL they don't know which code we have". A then prints out the list sent by colleague D and goes to discuss it with the colleague next to her office. She returns with two new numbers to search for which she jotted down in pencil – these numbers also don't fit. A explains: "This is our own list, the one the girl made for us, and it is already a little bit old". A looks at the data sheet – there are two part numbers, one for the front seat (21153) and one for the rear seat (211618), "when they buy it – when we sell it, we have to use the customer IDs for these parts". A calls L who asked for this change, tries another ID – maybe the customer AA has forgotten to approve these numbers – she calls L again who promises to contact AA about this – A makes a note on the print-out of the email message so as to remember.

Here we can identify several problems: Suppliers, in particular small ones, have problems in providing detailed information about their parts, often it is not clear whom to contact on the supplier site. That is why A has problems to retrieve this information. Different stakeholders use different part numbering systems and the mapping problems between these have not been resolved. These makes the retrieval of information about part id's almost impossible. There is no process support provided by KeyA's IT infrastructure for A: A switches between her email, different lists, including print-outs of lists, her post-its, phone conversations, printing out lists, carrying them to another office to discuss the problem, and so forth.

These two cases illustrate the problem of knowledge management at KeyA. Our approach tries to avoid such situations. In the next section, we introduce the concept of snapshots to show how to approach to this problem of knowledge management in distributed work environments.

3 The Theory of Knowledge Management with Snapshots

The concept of a snapshot is based on the network of projects, persons and artifacts. Persons work in projects and cooperate with other persons inside or outside the company. Persons create artifacts, exchange them with others they cooperate, modify or remove them. During a project, artifacts evolve. They host content, changes made, debates carried out, questions asked, answers given, etc. Looking into artifacts give us

[3] "A" is a material specialist at KeyA. The product, company and person names are blanked.

insight into processes connected to *content*. Additionally, we need the *meta-level information* like who accessed which artifact when, what was the modification, was there additional communication about these changes, who did work with whom in which project when, etc. Meta-level information is the information about people and their actions by using several artifacts and by cooperating internally in their company or with external partners. It helps construct relations between projects, persons and artifacts. It builds up a network that can be captured as a snapshot. If we apply this model in an enterprise cooperating with other enterprises worldwide, the picture becomes very complicated (Fig. 1). Persons from different enterprises share some artifacts. Some persons have direct exchange, some communicate with others implicitly by means of an artifact, some artifacts are exchanged automatically when workflows have been established.

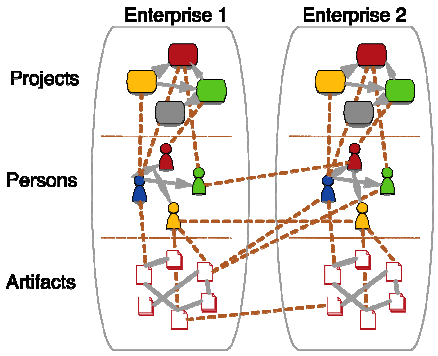

Fig. 1. Model of a snapshot consisting of a network of projects, persons and artifacts including content- and meta-level information in case of a cooperation between enterprises.

On the one hand, one can see the importance of artifacts in this model. On the other hand, the meta-level information needs to be captured and related to content-level information stored as artifacts. In our case study, artifacts we studied were mostly not related to each other electronically. Only persons using them knew more or less where to find which information about a certain project. If persons left the company, there was no possibility to recall past events and decisions at all. This can be solved by introducing the prototype *SUGGEST!* that we describe in the next section.

4. The Prototype: *SUGGEST!*

Based on the study at KeyA, we created an example network with real data that we used in our prototype called *SUGGEST!* The prototype is used to create, access and visualize snapshots in form of useful suggestions for users. With this prototype we try to show the principle underlying this approach. The prototype gives us the possibility to examine the model of snapshots. Unfortunately, we could not really introduce *SUGGEST!* in KeyA and see its impact on work practices around knowledge management (as described in section 2), because the project has already been finished

at the time of its availability. Still, *SUGGEST!* is the first prototype illustrating the application of a new concept, namely snapshots, in knowledge management.

SUGGEST! uses a data structure to store meta-level information and content (Fig. 2). Based on the entities of projects, people and artifacts, the data structure tries to link all entities to each other. The meta-level information excerpted from emails (Name=Subject, Content=Message Body, FromPerson, ToPerson), to-do lists (Name=Issue Name, Content=Issue Description, Deadline, Status, Responsible Person), documents' meta-information (Name=File Name, Content=Full text, Location, Type, Size) and of course also the explicit tagging and commenting are needed to create the context of data, to access the links between people, projects and artifacts in time.

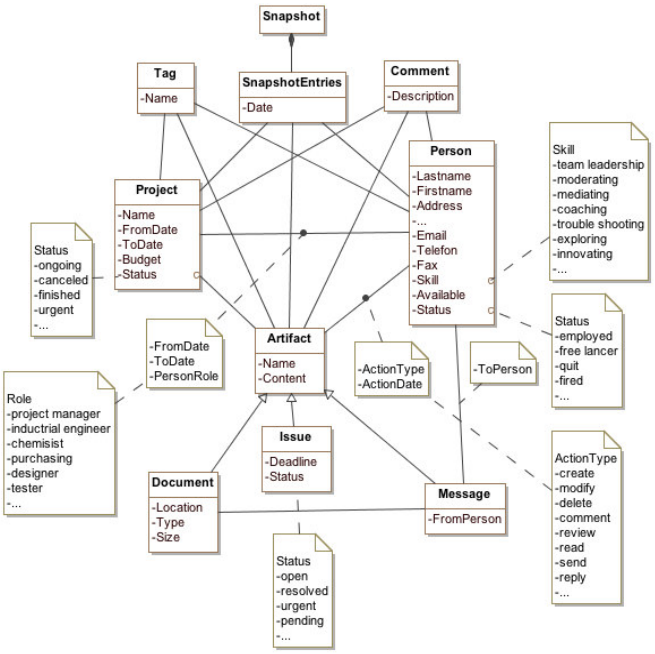

Fig. 2. Data model used in *SUGGEST!* Associations are meant to be many to many.

How are snapshots created? To feed the knowledge base with relevant content- and meta-level information, interfaces are needed. The meta-level information of emails containing the subject, from and to fields, the message body and the information about the attachments must be accessed by *SUGGEST!* to keep track of data exchange between project members and to identify email threads that enable to follow negotiations, problems or dense cooperation between people. To-do lists contain meta-level information as single issues, which needs to be accessed by *SUGGEST!* as well. The meta-level information of documents of any format is additionally needed to complete the picture. A file or configuration management system delivers the name of the file, its location, the name of the creator or modifier of the file and the date of modification.

It is obvious that interfaces must be able to gather all relevant information at the background without disturbing people in their daily work. Within an enterprise, the necessary access to mail server, file and configuration management server must be provided. It must be clear that the data collected this way will not be made accessible directly. It will just be used to create snapshots, saved in the *SUGGEST!*'s relational database, and to create suggestions. This is an important privacy issue. In case of collaborations with externals like suppliers, an explicit agreement must be set up to avoid confidentiality problems and legal issues.

How is this knowledge base used? *SUGGEST!* must enable data capturing, storing, indexing, useful retrieving and visualizing of meta-level information connected with relevant content-level information. Considering the usability criteria and building up on people's habits the following simple interface has been designed (Fig. 3). If someone wants to consume data managed in *SUGGEST!*, he or she needs to formulate the query in a natural language like in our case "who has experience with seat heating?". Depending on the meta-level information saved in the system, *SUGGEST!* responds with a list of search results (of person, project and artifact data) providing a context for the suggestion. If these first suggestions are not efficient or useful enough, more suggestions can be required. In our cases, for the first request a query like "which documents were emailed to customer C in project P?" would deliver artifact information, and for the second case a query like "who knows the part id of heating wires?" person and product information, and a query like "are there emails exchanged about part id's of heating wires?" artifact information as suggestions.

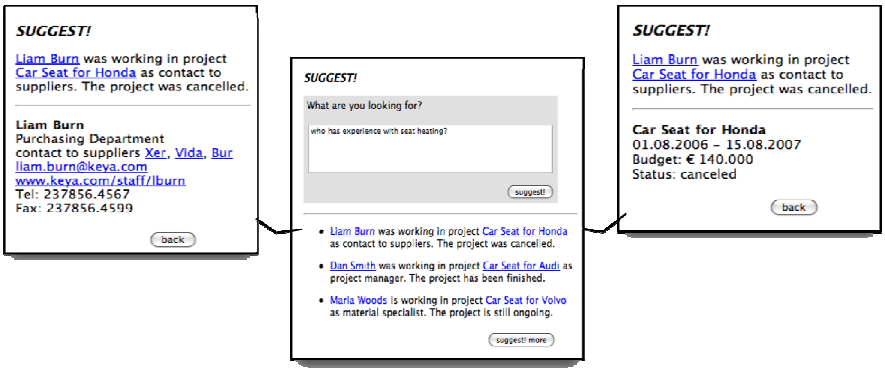

Fig. 3. An example query that delivers person and project data by *SUGGEST!Client*: The user request and the first list of suggestions matching the search criteria (middle), detailed information about person (left) and about project (right). User and project names have been changed. With "suggest! more" additional suggestions can be required.

5. Conclusions

This paper presents a new theory to knowledge management in organizations. It is based on the concept of snapshots that continuously capture properties of and relations between projects, people and artifacts in a cooperative distributed work environment. Based on an ethnographic case study, requirements to such systems are gathered and analyzed. A prototype called *SUGGEST!* has been developed to implement the idea of snapshots and to evaluate the functionality and usability of this knowledge base. *SUGGEST!* does not just collect all data available in an organization, it captures people's actions and exchange with others in concrete project teams. It is work in progress and needs to be developed further. It needs to be evaluated in a real work setting, as a running system and not as a prototype. More interfaces to legacy systems must be developed to build a wide range of possibilities for data capturing. In case of geographically distributed work, especially when externals like suppliers or third party companies are involved in projects, performance, integration, confidentiality and intellectual property issues must be addressed and solved. Still, we believe that snapshots are the correct approach for implementing a knowledge base for an organizational setting. We will continue on the development of *SUGGEST!* for further use and evaluation.

References

1. Agichtein, E.: Scaling information extraction to large document collections. In: IEEE Data Eng. Bull., vol. 28, no 4 (2005)
2. Chung, K.S.K., Hossain, L.: Network structure, position, ties and ICT use in distributed knowledge-intensive work. In: Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work, 545--554 (2008)
3. Georgalas, N.: A framework that uses repositories for information systems and knowledge integration. In: Proceedings of the 1999 IEEE Symposium on Application-Specific Systems and Software Engineering and Technology, ASSET, 128--135 (1999)
4. Klimathianakis, P., Loucopoulos, P.: DELOS - A repository based environment for developing network centric applications. In: Proceedings of the 9th International Conference CAiSE97, Barcelona, Spain, June (1997)
5. Penta, A., Picariello, A., Tanca, L.: Multimedia knowledge management using ontologies. In: Proceeding of the 2nd ACM Workshop on Multimedia Semantics, 24--31 (2008)
6. Sarawagi, S.: Information extraction. In: Foundations and Trends in Databases, vol. 2, no 1, (2008)
7. Senellart, P., Mittal, A., Muschick, D., Gilleron, R., Tommasi, M.: Automatic wrapper induction from hidden-web sources with domain knowledge. In: Proceeding of the 10th ACM Workshop on Web Information and Data Management, 9--16 (2008)
8. Speer, R., Krishnamurthy, J., Havasi, C., Smith, D., Lieberman, H., Arnold, K.: An interface for targeted collection of common sense knowledge using a mixture model. In: Proceedings of the 13th International Conference on Intelligent User Interfaces, 137--146 (2009)
9. Takahashi, K., Sugiyama, A., Shimomura, Y., Tateyama, T., Chiba, R., Yoshioka, M., Takeda, H.: Web-based knowledge database construction method for supporting design. In: Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services, 575--578 (2008)
10. Weikum, G.: Harvesting, searching, and ranking knowledge on the web: Invited talk. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, 3--4 (2009)