TechScreen: Networked Knowledge Management

Jürgen Dorn[1], Christoph Herzog[2] and Hannes Werthner[1],
[1]Institute of Software Technology and Interactive Systems, Vienna University of Technology,
Favoritenstrasse 9-11, A-1040 Wien, Austria,
email: {juergen.dorn|hannes.werthner}@ec.tuwien.ac.at}
[2]E-Commerce Competence Center, Donaustadtstraße 1, A-1220 Wien, Austria
email: christoph.herzog@ec3.at

Abstract

This paper describes a system supporting knowledge sharing in and between networked organizations. Knowledge resources are stored in a structured way, supporting the flexible association of access levels to knowledge resources and thus giving an organization the possibility to share knowledge while securing knowledge resources which are categorized as critical for competition. The system considers explicit knowledge stored in a Web-based information system and the implicit knowledge of human experts. Both types of knowledge are described by meta-data, in terms of an ontology and/or user created keywords/tags. These meta-data are used to find relevant knowledge sources, cluster these resources and to recommend certain resources to users. In this paper we introduce the system through a usage scenario at the university where students and university staff use the system to share knowledge.

1 Introduction

Technological knowledge on Internet and related technologies is rapidly growing. New standards by the W3C and other standardization bodies are published regularly, new open source components of frameworks are developed by Apache Software Foundation and others continuously and many software artefacts are provided in repositories such as the Free Software Directory of Free Software Foundation (http://directory.fsf.org/) or Sourceforge (http://sourceforge.net/). Also new releases of standard software by the large software vendors are offered in short periods. Many small companies and organizations struggle to be able to adequately use latest technologies demanded by their customers. Keeping up to date with open-source content management systems, AJAX (Asynchronous Javascript and XML), Web services or technologies for accessing Web pages by disabled persons are recent challenges for developers. Although there exist considerable knowledge on the Internet, usually there is not enough man-power to evaluate the different technologies and to decide which one to use in a certain environment.

TechScreen is a public funded project that tries to develop a community that supports the sharing of knowledge (especially experience) about Internet technology.

The following key questions are analysed in the project:

1. *How is the knowledge to be organized so that members can find available knowledge easily?*
   Ontologies are used to define a structure of different types of knowledge as well as to

organise the topics in which knowledge can be classified. Social tagging is used to organize knowledge in user-centered way, being better suited to deal with rapidly changing user demands. Moreover, not only knowledge explicitly stored in the information system but also implicit knowledge of the members of the community is considered. Members supplying certain content are considered experts in the respective domains and can be found and asked for advise by other members.

2. *How is the knowledge collected and interrelated?*
   We use automatic information extraction from available knowledge sources and employ text mining techniques to filter available knowledge from available user generated content and comments. This knowledge is structured by assigning content to concepts of an ontology and generating automated and user generated indexes. Knowledge can be associated with different confidentiality attributes allowing members to restrict the visibility to private, organizational or open access. Thus an organization may define that knowledge entered by their employees related to certain knowledge domains is confidential.
3. *How are users motivated to generate content and to share it with other members?*
   Users can define in a profile what their competencies, and what their professional interests are and which kind of information they provide should be visible by others. Todays knowledge workers should be interested to demonstrate their competencies. For this purpose a competence profile and a knowledge sharing profile of each member is computed from their entered knowledge. Furthermore, users can store knowledge so that they can document and access their own experiences quickly by different ways. By combining this knowledge with automatically generated content, additional hints become available. By enabling the access of other members to ones own contents these members may comment and discuss certain problems and, moreover, the user may achieve a certain reputation either in his own organization or also outside.
4. *How are companies and organizations motivated to share knowledge?*
   Today organizations have to be visible and to show where their competencies are. From the staff member competence profiles we compute a competence profile of the organization. Moreover, an organization has to achieve a critical mass of employees working on certain topics. Thus, an organization may derive from such a profile that new employees are to be hired to fill the gap. The access to knowledge assets of a company may be restricted so that core competencies of the organization are secured. On the other hand the organization should be able to document and show its core competencies.

One application scenario for TechScreen is a university with different faculties and institutes. Usually each institute has a small group of persons (normally technical personal) interested in Internet technologies and trying to build good information portals for students and staff members. In order to benefit from this knowledge, institutes need to share their experiences, a task where the TechScreen system will be successful if the access is easy and the system contains relevant knowledge. One possibility to generate an initial knowledge base is to access master and Ph.D. students working with these technologies. These students may be obliged to share their knowledge. In principle, the knowledge generated at a university should be publicly available and thus other organizations and private companies should be enabled to access this

knowledge. However, to achieve full benefits from such a knowledge base private companies should also participate. Their motivation should be to challenge the knowledge creation.

2 State of the Art

The term social software was created only recently, however, applications that follow this paradigm are much older. Due to different reasons there is some hype about these applications now. Thus new start-up companies offering such information systems achieve a very high financial rating through their large number of users and the large body of information. However, only a part of the social software applications are aimed at achieving a very high volume of users. Social software is also used to build smaller communities with a restricted access. Thus a company may invite its customers into such a community for online support on products and services of the company. Social software is also used to support knowledge exchange between employees of companies (Wenger 2004).

The success of social software depends on the motivation of its users. Either the number of participants must grow very large or users must be highly motivated to supply information into the system in order to motivate and attract other users. Communities are built around common interests. Often however, it is unclear what the common interests are and whether all community members share the same interests.

This problem is also investigated in knowledge management theory. A company should be interested that information is shared between its employees. Knowledge management systems have to be designed in such a way that individual members of the staff are motivated to share relevant knowledge through these systems. Davenport and Prusak (1998) describe three motivations that lead to successful knowledge sharing: reciprocity (if I submit something then also other community members are obliged to share information), reputation (if I submit much information I will be accepted as an expert) and altruism (I want to support this highly relevant community without any immediate benefits). These motivations are also valid for social software applications. A fourth motivation for participating in social software systems is the provision of special services for the participating user, such as data storage, email management and others. By offering such services an initial participation may be motivated.

For user acceptance of such systems the ease of use is also very important. Technological improvements such as simple interactive Web interfaces accessible with most Internet browsers lower the barrier to use such a system because users do not have to install software or to learn new interaction techniques.

Providers of community information systems must also get some revenue to provide services. The number of users is most important for financial evaluation of such companies. Google, for example, is the company having one of the largest user groups at the moment and thus achieves very high financial ratings. Thus, a provider is motivated to attract as many users as possible through good services. One of the most important services in the Internet seems to be the provision of relevant information (e.g the case of Google). And the easiest way to provide such information is to let the users create the content. Other options are automated creation of text by harvesting the Web with crawlers, simple indexing of existing resources by search engines and knowledge extraction through text mining techniques.

2.1 Knowledge Indexing

Social or collaborative tagging is a process where a number of users (the members of a community) assign tags (key words) to certain information or knowledge resources. Such resources can be, among others, photos (e.g., Flickr), videos (e.g., YouTube) or URLs (e.g., Del.icio.us). The set of tags used in a community is called a folksonomy. A folksonomy is formally a hypergraph containing three types of nodes related by assignments. A folksonomy F = (U, T, R, Y) is a quadruple where U is the set of users, T the set of tags, R the set of resources and Y the set of assignments where one user relates one resource with a tag. Often tag sets are used to characterize one resource. Although users are free to use such tags, investigations have shown that at least in large communities some convergence in using common tags can be seen (Golder and Hubermann 2006). Additional means to improve this convergence are proposed in order to support further reasoning processes. Mainly the efforts towards this end result in some kind of feedback to the user showing which tags exist in the system and which are the most popular ones. A further step towards improving this meta-data is to apply ontologies defined by experts. In this case either users can assign terms defined in the ontology to a resource (Herzog et al. 2007) or certain concepts of an ontology are automatically assigned to resources by means of analysing the resource with text mining algorithms.

With a certain grade of reuse (using same tags, tagging same resources and users that regularly tag resources), social network analysis methods can be used to detect interests of users, to find clusters of users with similar interests and to recommend resources to users (Schmitz et al. 2007). For some investigations (e.g. trends in interests) the time when the tagging has taken place must be stored additionally.

2.2 Competence Management

Competence management is a recent trend in enterprises to better organize the development and recruitment of the work staff. Today, a fine-grained modelling of available knowledge and capabilities of employees is required to plan knowledge-intensive projects in companies. We developed an information system to model student competencies supporting the explicit planning of student's competencies (Dorn and Pichlmair 2007). Thus a student may compare his or her actual competencies with a predefined goal profile to determine which lectures, projects or other experience would help to reach the goal competencies. Especially the development of soft skills is only seldom considered in a university curriculum today. We focus on documenting and assessing such competencies in existing courses. Students' competency profiles can also be used to support the recruiting in companies if these companies have comparable competence management systems. Moreover, such competence management can support the staffing in inter-organizational projects (Dorn 2006). The competencies are defined through a formal ontology. For the computer science and the information system curriculum we have defined approximately 140 different competencies. For example, Web Engineering is a competency with a knowledge and an experience aspect. Thus a student may achieve a certain degree of knowledge in a lecture and by completing a project in this domain he or she may also achieve a certain grade of experience. However, from this measurement alone we do not know whether the student has experience in managing a certain application server. Since the content of such university courses changes very frequently and systems are generated continuously, an ontology cannot always represent the latest state of fine-grained competencies. The use of tags can help to supplement the ontology

meta-data with a highly dynamic and easily usable vocabulary. While lacking synonym control and explicit semantics, tags can be used to indicate experience with specific technologies belonging to the area of a defined competence (e.g, indicating which specific application servers the student has experience with in the example above).

3 Design

The TechScreen community management system is designed with a focus on a flexible integration both from a user/organizational perspective and from a technical perspective. Individual organizations can use the system as a stand-alone service as well as a tool to share knowledge between organizations. From a technological view, we enable the flexible integration of new components and technologies for knowledge management. In the following we first describe the organization perspective, then the technological one and finally the approach to competence management followed in the system.

3.1 Management of Users and Organizations

The objective is to attract users and to encourage them to supply information and to maintain their profile. The incentive for registering with the system will be achieved by providing better services to registered users. For a registration at least a legal email address, a nickname and a password are required. Supplying phone numbers and other contact information is encouraged to support communication outside of the system. The name of one or more organizations with which a user is associated, can be supplied, but to achieve access to information restricted to a certain organization, the organization's administrator has to approve this affiliation. A user can specify interests and competence fields in the profile.

The following services are available for registered users:

1. A user can specify filters in his profile to indicate which information he is interested in and how he wants to be alerted about new relevant content.
2. A user can generate a competence profile showing his competencies based on his profile specifications, submissions, read articles and applied tags. Thus other users may find this user as expert in certain domains.
3. The user may specify in to which extend and in which form he is willing to share his knowledge and expertise.

The system will support the participation of organisations as an entity. On one side the system may be used as a knowledge management and sharing tool inside of an organization to find experts in certain (technical) domains and to find information about certain topics. Moreover, organizations should be motivated to share such knowledge with other organizations and their members. Organizations may also pull knowledge out of the system by posing challenges to the community.

The following services should motivate organization to participate:

1. A member of an organization may find a competent person in a certain competence field outside of the organization.
2. An organization may generate competence profiles generated from submissions of their members and from explicitly mentioned competences.

3. An organization may pose problem statements in the system that are solved by the community.

An organization may be represented with sub organizations, therefore complex management and delegation of privileges are supported. Each organization needs an administrator which decides on whether certain users are members of the organization.

Knowledge affiliations are agreements between organizations to share knowledge. Two styles of agreements exist: transitive and non-transitive affiliations. If organization *A* is willing to share its knowledge transitively with organization *B* and organization *B* has an agreement with organization *C*, then *C* has also has access to *B*'*s* knowledge.

Knowledge entered by a person being member of an organization can be classified into four confidentiality levels

1. *Open* means that the knowledge resource can be seen by every user entering the platform (also unregistered users)
2. Corporate level means that all members of affiliated organization may see the knowledge resource.
3. *Restricted* level restricts the visibility to the organization and
4. *Private* level means that only the user may see his entries.

An administrator of an organization can assign confidentiality levels to certain types of knowledge and information resources. Thus each field of a user profile can be classified individually. Knowledge and information resources are classified in different types as described in the next section. Each of these resources may get an organization wide confidentiality level, assuring that certain information and knowledge is not shared publicly.

For certain knowledge domains dedicated members can become moderators. These are experts in the domain and are especially motivated to develop the knowledge in that domain.

3.2 Knowledge and Information Resources

The system stores explicit information/knowledge in different resource types. Moreover, humans and organizations working with the system are interpreted as resources with implicit knowledge. A search for knowledge may return an explicit resource as well as link to a person/organization that has the required knowledge or expertise. The management of this "implicit" knowledge is described later. The explicit knowledge can take two forms: it may be some content in the system or a reference to a knowledge artefact outside of the system.

The resource types are stored in different "templates" defined by an ontology. The root element "knowledge resource" has some basic attributes inherited by all children. A knowledge resource has a "created" time attribute and may have a "valid-until" attribute describing the life-time of a resource. Each knowledge resource has a "creator"; either a user of the system or a software agent that has entered the resource into the system. The creator may define an individual confidentiality level to the object. Per default a resource is visible by all members. If the creator's organization has set a stronger confidentiality level, then this level is the default. However, the creator may assign an individual level to the object.

The first child of a knowledge resource is a "challenge" object. An instance of such an object contains individual problem descriptions such as for example "setting up a directory service on a server". There may be complex challenges composed of smaller problems. Further, we distinguish whether such a challenge needs a new solution or a simple adaptation

of an existing solution which is determined by the community. The next child of the root element is a "solution". Challenges and solutions may be assigned to each other. Thus there may exist different solutions for a problem and a solution may be applied to different problems. A solution can consist of different steps which again can be solutions to smaller problems.

A third child of the "knowledge resource" concept refers to documents containing information used in problem solutions. Solutions and solution steps may reference these documents. We distinguish scientific articles, standards, manuals, tutorials, white papers and slide copies. These documents are stored in the system in PDF-format. The concept "Software", with further children, is also a "knowledge resource" having certain attributes and references to download locations. Educational events disseminating knowledge related to domains in the TechScreen system is a further child concept. The last child concept is "links" which can be used to reference further knowledge outside of the system. Here again different child nodes are defined for blogs, wikis, forums and more.

Besides this knowledge resource type ontology, a second domain ontology is defined. In our first system this ontology describes Internet related concepts such as "Web server", "XPath" or "DNS". We could exchange this ontology to share knowledge in a different domain.

3.3 Tagging

Each knowledge resource entered into the system obtains meta-data describing the resource. Some descriptions such as the creation time or the creator are assigned automatically. By assigning a resource to a certain concept of the ontology, further tags associated with that concept are also entered automatically. But in addition to that a user can assign individual terms characterizing the resource. Each of these terms is a quadruple resource/user/tag/time stored in a database. As already mentioned these user supplied tags serve as a secondary meta-data description of the knowledge resource, allowing for a more detailed specification of the stored knowledge. Since the tags of a folksonomy constitute a highly dynamic and user-centred vocabulary, they can succeed in describing complex and rapidly changing domain details (like individual technologies). While they lack explicit relations and semantic expressiveness, they provide users with an emergent vocabulary that is best suited for the users' needs at a given time.

3.4 Knowledge Extraction and Integration

Many knowledge resources already exist on the Internet. We use standardized methods to extract this knowledge into our system using the Lixto technology (Gottlob et al. 2004), combining and compiling identified knowledge items into comprehensive PDF documents (e.g., containing an installation tutorial, an FAQ and forum entries concerned with a specific application server). With text mining methods we extract keywords from these resources to add these as resource tags. A simple solution is tagthenet (http://tagthe.net/). Improved solutions would use a domain vocabulary or also make use of our domain ontology. Furthermore, text mining is also used to identify relevant blog entries and forum threads for a specific topic.

3.5 Competence Evaluation

In the system the competence of members and of organizations shall be measured. The vision is that the system will be integrated with the competence management system described by Dorn and Pichlmair (2006). However, at the moment only experiments how far the measurement of competence based on the participation in the system is evaluated. In the existing competence management system consolidated competences are measured (e.g. Project project management competencies) by a supervisor and peers (Dorn et al. 2008). In TechScreen here we want to measure competencies in technologies that evolve rather fast.

The system tries to evaluate each submission (an article written, a comment on another resource, a tagging as well as reading and evaluating another resource) of a community member. Taggings of the resources are potential competencies. If a number of community members have tagged resources with the term "application server", then this becomes a certain fine-grained competency. A member may have submitted different resources having assigned such a tag. Then we compute an aggregated value from these resources assuming that the member has competency in "application servers". Different types of resources have different influence on the computed value and moreover, the evaluation of a resource by another person can change the worth of one's own submission. For example, someone may have rated an article as very good. Each submission may have influence on different competence areas. Since we are still experiment with the aggregation algorithm, we give here no further details.

Based on all submissions we may now compute a profile describing the competence in different areas. This profile can be used to find persons having certain competencies.

The profile of an organization is computed by an aggregation of the competencies of its members.

4 Prototype

The first prototype is developed for knowledge sharing in a university environment. The addressed users are students writing their master or Ph.D. thesis on a topic related to Internet-based applications. These students have achieved a basic education in this domain and now have to solve specific problems in this field of technology. Here we consider regulations existing at the computer science faculty of the Vienna University of Technology. However, similar procedures also exist at other universities in Vienna and abroad. In Vienna there are also other universities with similar curricula whose students can easily be integrated. In the following we describe only the knowledge sharing process of master theses.

A master student has to find a problem that can be solved in a master thesis. This may be a problem coming from a company where the student works, a problem formulated by a professor at the university or some problem the student is interested in due to some other reason. The student is asked by the faculty to describe this problem. This problem description, together with administrative data of student and supervisor, can now be entered into the TechScreen system. The document is uploaded and some meta-data are automatically created by the system. These are derived from the implicit process model. Thus the document is characterized by "problem", "thesis" and "date". Further, the student has to classify it into the existing domain ontology by referencing one or more concepts of the ontology. The system analyses the text with a text mining algorithm and proposes a more fine grained meta-data description using ontology elements and also tags, where the student may use an own

vocabulary of tags to extend the meta-data. The second document a student has to upload is a first presentation to be given in the so-called "Diplomandenseminar". This document is classified as presentation and meta-data can be inherited from the first document. Moreover, the actual presentation is entered as an event with further meta-data. After completing their work, students submit the final thesis with an abstract, a final presentation and a poster presenting the results for a more general audience. An additional document is the review report of the supervisor. This of course is a confident document, but it is an important feedback for the worth of submissions of the student.

Typically, a thesis contains a state-of-the-art chapter. This chapter contains descriptions of methods, techniques and systems that can either be reused for solving the problem or are comparable problem solving approaches. If the system is fully developed by the users, the student will find many links, texts and documents in the system which already describe these artefacts. However, in the starting phase of the TechScreen system, the student will mainly find automatically generated content. The student is now forced to enter state-of-the-art artefacts himself. If he or she finds descriptions of other students, he or she is asked to critically reflect on these descriptions and to give comments.

Each input (link, document upload, text, tagging and comment) is accounted and a profile is generated for a student showing her/his performance (number of submissions, time periods, …) with the possibility to compare it with other students. A supervisor having access to the system can now see whether the student is willing to share knowledge. Furthermore, a competence profile is computed showing the interests and competencies of the student. This contains on the one hand the competencies related to the content of the thesis and on the other hand social competencies derived from the behaviour in the system.

The student's profile can be partly seen by other students. Thus, one student may contact other students in the TechScreen system to request appointments in order to discuss certain issues the other student should be competent in.

As already mentioned, a supervisor has to review a student's thesis. TechScreen can support this review process. By giving statistics to the supervisor some basic evaluation is possible. The supervisor can evaluate other aspects in a structured template assigning reviews to students' submissions. A good review will lead to an improved visibility of the student's competencies. Moreover, supervisors are potential moderators of certain knowledge domains.

5 Business Models

In the TechScreen project a first prototype was developed for sharing knowledge on Internet related technologies. An application of the system in a different domain would require another domain ontology. Although the main system shall remain an open knowledge sharing tool where any interested person may participate, a revised system could be applied in a more "commercial" version. The software could be used as a standalone application in organizations with own servers to support the knowledge sharing inside of such an organization. This means the organization buys the software and the management has full control on all aspects of the system. Since the storage of information, the computation of different evaluations and the processes are standardized within the TechScreen system, organizations having such a system, can also exchange knowledge with each other.

Especially for smaller companies that do not have own IT units the outsourcing of such a service on a TechScreen server would be more promising. Here common services and premier services are distinguished. If the organization wants to use premier services such as hiding

knowledge or special evaluations, these may sold to the organization separately. It would also be possible to integrate specific existing online forums to such a system.

The system could also be opened for advertising certain products related to the content of the system. Members should be able to filter out commercial advertisements or at least should easily recognize the difference between commercial ads and user generated content. A company developing IT tools could also use such a community platform to communicate with its customers and to enable its customers to build up a community.

The system could also be used to offer consulting capabilities of single persons or organizations. This can again be implemented as a paid service.

6 Implementation

The first prototype was implemented with an Apache Web server, MySQL database management system, PHP and Drupal 6.1 as content management system. Further an LDAP server is used for authorization and user management and Protégé with OWL is used for the ontology.

7 Discussion

We have described a community information system supporting the exchange of knowledge between community members and organizations. We distinguish knowledge that can explicitly stored in such a system and knowledge that exist only implicitly in the head of the human members or in the organizational memory. The indexing of the knowledge is by automated techniques, user-initiated tagging and by assignment to concepts of ontologies. These indexing techniques support also the finding of knowledge resources, i.e. human experts, organizations or simple documents.

There exist several ways and motivations how organizations can share knowledge with each other. Such a knowledge exchange of course must have benefits for all partners. However, these benefits may be of different kind. Most appropriate is always that knowledge is exchanged with other knowledge and we try to evaluate also the value of such knowledge. Experiments are still on the way and no final conclusion about this can be drawn.

8 References

Davenport, Th. H. and Prusak, L. (1998) Working Knowledge – How Organizations Manage What They Know, Harvard Business School Press.

Dorn, J. (2006) Competence Management in Virtual Enterprises, in Proceedings of XML-Tage, Berlin.

Dorn, J. and Pichlmair, M. (2007) A Competence Management system for Universities, in Proceedings of European Conference on Information Systems, St. Gallen, pp. 759-770.

Dorn, J. and M. Pichlmair and K. Schimper and H. Tellioglu (2008) Supporting Competence Management in Software Projects, in Proceedings of International Conference on Concurrent Enterprising, Lisboa

Golder, S. and Huberman, B. A. (2006) Usage Pattern of Collaborative Tagging Systems, Journal of Information Systems 32(2) 198-208

Gottlob, G., Koch, C., Baumgartner, R. Herzog M. and Flesca, S. (2004), The Lixto data extraction project – back and forth between theory and practice, in Proceedings of the Symposium on Principles of Database Systems (PODS-04)

Herzog, Ch. , Luger, M. and Herzog, M., (2007) Combining Social and Semantic Metadata for Search in a Document Repository in Proceedings of Bridging the Gap between Semantic Web and Web 2.0, International Workshop at the 4th European Semantic Web Conference, June 7, 2007 - Innsbruck, Austria.

Macgregor, G. and McCulloch, E. (2006) Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool, Library Review, Band 55, Nummer 5

Wenger, E. (2004). Communities of practice: learning, meaning, and identity. Cambridge et al.: Cambridge University Press.

Schmitz, Ch., Grahl, M., Hotho, A., Stumme, G. Cattuto, C., Baldassarri, A., Loreto, V. and Servedio, V. (2007) Network Properties of Folksonomies, Proceedings of the WWW 2007, Banff, Canada.