Implicit Social Network Construction and Expert User Determination in Web Portals

Andreas Nauerz
IBM Research and Development
Schoenaicher Str. 220, 71032 Boeblingen, Germany
andreas.nauerz@de.ibm.com

Georg Groh
Technical University Muenchen
Boltzmannstr. 3, 85747 Garching, Germany
grohg@in.tum.de

Abstract

Especially within the context of the ongoing Web 2.0 hype, social software has recently gained more and more momentum. Social software is software that focuses on the interaction and collaboration between users being part of social networks. So far, most of these networks are constructed explicitly by users themselves. On the other hand Web Portals give users a central point of access to relevant information. They provide users with a highly collaborative environment, but, until now, required the users that interact and collaborate to know each other. As part of this work we present a system additionally able to construct social networks implicitly by observing users behavior in Web Portals. We aim to use the derived social networks to enhance interaction and collaboration within the community.

Introduction

Recognizing the value of internet technologies, companies started to establish intranets to provide their employees with a central point of access to enterprise specific information. Initially, intranets focused on presenting the most valuable and widely used information to employees providing them with quick and efficient information access. But the amount of information accessible via those intranets quickly grew and finding the right information became more complex and time consuming.

Hence, many companies successfully introduced Portals to manage the upcoming information overflow. Portals provide users with a central point of access in a highly personalized manner. They allow providing information and applications based on the role of a user within the enterprise. Understanding the role of a given user and his information needs allows efficient targeting of the content. Today Portals are comprised of a huge amount of content. They often grow tremendously and it again becomes more time consuming to find and access the right information throughout a particular context.

Usually single users use only a small subset of all the content available regularly. The services based on this content provide them with means to accomplish typical tasks according to their job role. They can be regarded as experts with respect to this content. However, users sometimes need to access content they are less familiar with. For example users traveling seldom nevertheless have to access and use Portal pages allowing them to book flights, hotels and cars and to do their travel expense from time to time. As they are less experienced with dealing with rarely accessed content they often do not know how to use special services in connection with that content or, even worse, do not even find the content or the respective services at all.

In these scenarios users may be searching for experts which could help out. But most often teams are comprised of users having similar roles and hence finding experts by searching in the user's own team (or immediate social environment) might not be successful and the real experts might remain unknown. Generally, the suitability of a person to act as an expert depends on a variety of factors such as skills, availability, access to appropriate media, and cohesiveness with other participants. That means that the problem of selecting suitable users is challenging because of people being globally located in modern enterprises not directly knowing each other. Thus, the question is how to bring users and experts in touch.

In the following we will present a system able to determine which users tend to be experts for which content and able to detect when users access content they are less experienced with and in which they have already got lost. The system dynamically provides the less experienced users with options to get in touch with expert users.

In addition, Portals, especially Enterprise Information Portals, provide their users with highly collaborative

environments. Solutions like IBM's Lotus Quickr [15] are even sold as virtual teamspace solutions. They allow virtual teams to be set up on-demand and provide team managers with options to explicitly create virtual teams. Typical functions allow team members to interact and collaborate by storing and retrieving information in wikis, blogs, by storing, retrieving, and sharing documents as part of document repositories, by jointly executing workflows stored in underlying workflow engines and so forth.

However, these teams are created explicitly by some team manager. Similarly, in other systems (e.g. LinkedIn [16], Xing [17], …), social networks are constructed explicitly by users themselves. The interconnection between them is explicitly triggered by one user and accepted by another. Only users already knowing each other can get in contact this way.

Obviously, in large enterprises users might have similar job roles or might at least perform similar tasks often without being in the same team or even knowing each other. In addition to be able to suggest expert users to others, the system we will present detects users behaving similar and provides them with means to get to know each other.

In the remainder of this paper we describe some related work, the concepts we use to construct (implicit) social networks and to determine experts and, finally, the implementation incorporated into IBM's WebSphere Portal [18] providing users with access to explicit contacts as well as implicit (also referred to as potential) contacts, contextual contacts and expert users.

Related Work

Generally, systems that help to find experts are called expertise finders or expertise location engines [1]. A general architecture for recommendation systems that allow locating experts is described in [2]. More specifically Streeter et al. present *Who Knows*, a system which recommends experts having knowledge in specific topics based on profiles created from observing the documents they have selected and worked with previously [3].

Newer systems that use information about (either explicitly or implicitly constructed) social networks to find experts have also been explored. In [4] Kautz et al. present *ReferralWeb*, a system where people are referred to other people based on relationships between people identified through co-authorship. *ReferralWeb* is an interactive tool on the web that helps people find short referral chains between themselves and experts within a certain area. It uses publicly available web pages to create the referral chain. No additional information needs to be entered by the users of the system. Similar systems are described in [5] and [6]. In [7] Yenta et al. present a system where profiles are created based on the observation of personal data (files in the file system, newsgroup articles, e-mail messages etc.) and actions being performed. These profiles are then leveraged to determine users' expertise and to recommend them as experts to other users.

A system analyzing users' tagging behavior to find experts is presented in [8].

From an algorithmic perspective, an evaluation of available search strategies to find experts in social networks is performed in [9].

Our focus lies on the application of three different metrics, namely *web (usage) mining*, *tagging behavior analysis* and *social network analysis* to (implicitly) construct social networks and recommend expert users. Especially web usage mining is a favored mechanism to analyze user behavior in Portal systems, but has rarely been used to construct implicit social networks. We regard the combination of the three mentioned metrics as promising. Moreover, our system represents a specific solution for (Enterprise Information) Portals.

Expert User Determination and Implicit Social Network Construction

A Web Portal is a site (often a system) that creates a single point of access to information (collected from different sources) and services in an often highly personalized manner. Figure 1 shows the general architecture of such Portals and of our system:

Within Portals the actual content is provided by hierarchically organized pages which can be accessed via navigation bars. Pages are equipped with portlets which provide dedicated information and services to users.

To be able to implicitly construct social networks and to recommend expert users to others requires understanding of users' behavior, their interests, preferences and knowledge. Therefore we construct user models reflecting a user's behavior. We use static information from users' profiles (describing their age, native language, etc.), as well as dynamical information which we retrieve via web usage mining, by analyzing users' tagging behavior. As an addition to that, explicit social networks that are constructed by the users themselves are also taken into account.

The social network visualization component uses the information stored in the constructed user models and from the social network representation to visualize the latter and to recommend expert users.

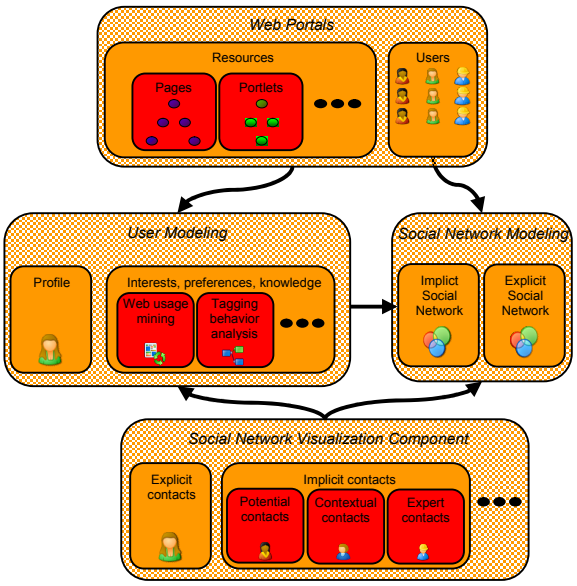

Figure 1: System architecture

User Modeling

As mentioned above, we are focusing on three metrics to create models reflecting users' behavior within the Portal: *web (usage) mining, tagging behavior analysis*, and *social network analysis*. These user models are used to determine the expertise of users with respect to certain (content) areas of the Portal and, by that, to implicitly construct social networks.

According to [10], *Web mining* is the application of data mining techniques to discover (usage)-patterns within web data in order to better understand and serve the needs of users of web-based applications. Especially web usage mining can be viewed as the extraction of usage patterns from access log data modeling certain aspects of the behavior of users.
Our system has been incorporated into IBM's WebSphere Portal. Its analytics engine logs and reveals information about, among others, the following events:

- Page management
  (creating, reading, updating, deleting a page)
- Requests of a certain page by users
  (including contained portlets)
- Session activities
  (login, logout, timed out, login failed)
- User management actions
  (creating, reading, updating, deleting users and groups)

That means that analyzing the log the Portal analytics engine creates allows us to understand which pages and portlets a user is typically working with. The first assumption is that users working with certain pages and portlets more often have more expertise about how to use them than other users have. The second assumption is that users that typically work similarly (e.g. with the same pages and portlets) should be modeled with a stronger relation within the social network we construct. For example, if users A, B, and C often work with the pages and portlets underneath the page entitled *social computing* we can, on the one hand assume that they have knowledge about how to deal with the pages and portlets provided here, and, on the other hand assume that their relation within the social network must be strong as they do similar things and might be part of the same team. A user D accessing the same pages and portlets rarely can then be presented with A, B, and C as experts when dealing with the information and services provided, especially if the system has recognized that he got lost. Latter can be recognized, e.g. if he is navigating in "circles".
Obviously the user model must allow us to calculate the utilization of pages and portlets from the historical data available. We do this by measuring how often a user interacts with certain pages and portlets. Of course we must be careful as not every click to a page means that the user is really working with it – it might just be a page the user clicked to be able to reach another page in which he is really interested. In the following, we say that every time the user accesses a page the page receives a hit. Every time the user really interacts with a page (i.e. uses a portlet on the page or stays longer on the page than he is averagely staying on pages) to perform a task or to receive some information the page receives a target hit. I.e. that if a user navigates from a page *S1* to a target page *Sn*, *n-1* pages receive a hit but only one page (*Sn*) receives a target hit. Hence, to calculate the utilization only the target hits are of interest. Since for some pages no interaction is taking place at all, e.g., because they are comprised of portlets only displaying some kind of information (e.g. weather portlets etc.), we also measure the duration the user visits pages and assume that the page received a target hit, if the duration was longer as the mean duration with which users visit pages.
More generally, we apply techniques from the area of frequent set mining [10] to analyze the usage of pages and portlets. We use algorithms like Apriori [13], a standard association rule mining algorithm, to determine items, such as pages and portlets that co-occur frequently between different users to find those users behaving similarly. We apply algorithms like GSP [14], a standard sequential pattern mining algorithm, to determine sequences of items, such as pages and portlets, that co-occur frequently between different users to find those users traveling similar paths often and hence behave similarly.

Besides applying web usage mining techniques we analyze the *tagging behavior* users' show to determine their interests and preferences. The tag engine we have incorporated into the Portal system allows users to annotate uniquely identifiable resources of the Portal such as pages, portlets and even other users. The general assumption is that tagging expresses interest in a resource. Hence, we can assume that users tagging certain resources often have

knowledge about how to deal with them. Tagging pages and portlets expresses knowledge about how to use the services provided by them, whereas tagging users expresses a relation to them. Moreover, tagging users might provide us with insights about their expertise. For example, if a user A tags a user B with the term *social-computing* we can assume that user B has knowledge about social computing. If other users have already tagged other resources such as pages and portlets with the same term this can be regarded an indication for user B being an expert in how to deal with these resources. Thus, analysis of the tagging behavior allows us the refine the previously constructed user models.
Furthermore, analyzing and comparing the tagging behavior between users allows for partitioning them into groups of „similar behavior“ which provides us with means to refine the implicit social network, too. In [11] algorithmic techniques are discussed for detecting groups of users on the basis of interests.

Finally, the analysis of users' explicit contacts allows us to determine users' interests and preferences, too. The assumption is that the fact that users directly know each other can be an indication for similar job roles and hence for sharing similar knowledge.

We additionally extended some out-of-the-box portlets that the Portal system comes with. Especially communication-centric portlets like e.g. wiki-, blog- and mail- portlets provide us with even more insights into user expertise and, in addition, with further means to implicitly construct social networks between the users of Portal systems. E.g. from a communication content perspective, if users respond to threads on a certain topic it can be concluded that they demonstrate knowledge with respect to that topic.

Social Network Modeling

From a social network perspective what we get is an explicit social network, determined by having the user establish explicit social relations, and an implicit social network derived from web-usage mining and user interaction with the portal in general (e.g. tagging) through a set of heuristics. The heuristics will never be 100% accurate on the one hand and we cannot expect the user to explicitly establish or rate the relations to other users in all cases on the other hand. Furthermore, the explicit ratings have the disadvantage that they are often limited to the user's immediate social environment which is the first address when looking for help anyway. "Horizon broadening" social recommendations [12] can only be achieved through the incorporation of implicitly derived relations. So what has to be done is to combine the explicit and the implicit social networks to yield a more reliable result.
For example in [11] approaches on how explicitly stated group models can be combined with implicitly derived group models are developed. For social networks, comparable approaches can be applied. First of all, we can assign every relation in the combined network a reliability weight. The most straightforward approach is summing up: Having n strength weights for a social relation implicitly derived with the help of n heuristics and m explicitly stated strength weights for the same social relation we can simply sum up the $n+m$ weights to yield the weight for the combined network. This will (from a single user's perspective) mainly lead to strengthening (first grade) explicit relations with users which usually or at the present moment are dealing with the same topics. Those are (for reasons stated above) not particularly valuable with respect to new "horizon broadening" expert recommendations. But since all users of the portal have their own star-shaped social networks combined into an overall social network, the approach of combining the weights of explicit and implicit relations also leads to an increase in trust (with respect to expertise relevance) into paths of *length=2* to users which the active user does not know directly. This is because the expertise relevance of the two respective direct relations involved is expected to be increased through the summing up of weights and thus the justification to transitively deduce expertise relevance is also increased. Thus we can incorporate the weights of *paths≥2* from user u_i to user u_j into the strength of the expertise relatedness between u_i and u_j (e.g. by weighting inversely with the length of the path).

Social Network Visualization Component

As the result of the analysis described above, we are able to provide the user with context-dependent information about his explicit and implicit social networks. In particular, we can provide the users with experts, e.g., when he gets lost. In this section, we describe the component visualizing social networks in our system. This component been realized as a specialized portlet (Fig. 2) that can be accessed on every page (on demand).

Structure of the Social Network Visualization Component

At the top of the portlet contacts are listed which have explicitly been added as such by the user. Adding contacts explicitly demands sending a request to the contact to be added which, in turn, can accept or reject this request. While users explicitly add other users as contacts the explicitly modeled part of the social network is constructed.
The next section displays contacts the system has determined to behave similarly (based on the web usage mining etc. performed) or might be part of the same or closely related teams (based of the analysis of the entire social network) to the current user. These are contacts the user might want to get in touch with, generally to share knowledge. E.g. users reading the same papers, attending the same conferences, communicating with a lot of the same contacts might want to start collaborating on a certain topic or at least share some ideas.

The third section displays contacts currently "performing similar actions" within the Portal. In contrast to the contacts being displayed in the previous section, the contacts being displayed here do not necessarily have to behave generally similar to the user, but can be regarded to behave similar in the current situation (with respect to the current context). Users might be listed here because they are on the same page, use the same portlet and so forth.
The last section at the bottom of the portlet displays contacts the system has determined to be experts with respect to the (content) area (again, based on the web usage mining etc. performed) currently being visited by the user.
The list of users displayed in the last three sections dynamically changes as the current user interacts with and navigates through the Portal, while the first section always displays a static list of contacts.

Functions of the Social Network Visualization Component

There are some typical functions that the current user can invoke when working with contacts displayed as part of the portlet. Most of these functions can be accessed from a context menu that opens when a contact is clicked.

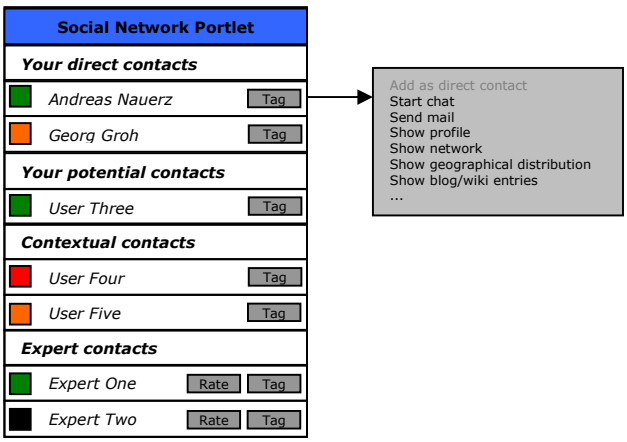

Figure 2: Social Network Visualization Component

People-awareness. To ensure that users contact other users being available more likely than the ones being unavailable and to ensure that users do not disturb latter, e.g. during meetings, an online status is displayed for each single user displayed as part of the portlet. Here, green indicates available, orange away, red do not disturb, and black offline.

Tagging and rating. Users being displayed as part of the portlet can be tagged, expert users even be rated. Tagging, the association of keywords to resources (here users) can allow for a user-driven categorization of other users in an autonomous way. Additionally, rating allows assessing how helpful an expert was in a concrete situation or the respective expert's general degree of expertise.

Tags being assigned to users can be evaluated by our system to determine in which area the specific user might be an expert in. Obviously, because of synonyms and polsemy, normalization has to be performed.
Furthermore, ratings can be used as a metric to decide which subset of all available expert users with respect to a certain topic should finally be displayed.

Profiles. The "profiles" function allows for accessing the profile of a contact. It displays information about the contact's official job role, his position within the organization's hierarchy, geographical location, post address, mail address, phone- and fax numbers, wiki- and blog entries being posted by the contact and for which areas of the Portal the system has determined him as an expert. The profile functions can be regarded as an enhanced address book. Especially for the non-explicitly added contacts this function might be of importance to the user. It allows him to get more information about the contacts the system displays and hence to understand why these are displayed at all. It is also possible to correct the system's guesses as a learning feedback. The information retrievable from the profile might form a basis to decide whether an implicitly added contact should be added as an explicit contact or not.
Generally, all implicitly added contacts can be added as explicit contacts by choosing the corresponding function from the context menu.

Social network visualization. This function displays the user a visual depiction of how his (explicit and implicit) contacts are related to him and to each other. It presents a graph which nodes are the user's contacts.
It allows users to determine which users are part of the same team, who personally knows whom and hence allows e.g. to find backups for persons being currently unavailable easily. Implicit contacts are displayed in a different color than explicit ones making it easier to classify them and to decide whether to make them explicit contacts or not.

Geographical distribution. This function displays the contacts being displayed as part of the portlet as marks on a GoogleMap. We regard this as an important features as users tend to preferably get in touch with contacts located nearby.

Remote assistance. Especially in cases where a user is seeking for help remote assistance is a highly appreciated function. Users can contact an expert and ask for remote assistance. The expert users can than take over control of the users' system and assist while navigating through the Portal system and while using various portlets.

Contacting, exporting and importing options. Of course, to get in touch with other users options to start instant-messaging or to directly send a mail (e.g. if the user to be contacted is offline) are provided. Contact's profiles can be

exported into various formats, e.g. as vCards to import the data into other programs (e.g. mail clients).

Issues

Issues are mainly related to privacy aspects. Monitoring any kind of user activity is a sensitive issue and any system that monitors and makes inferences based on user activity should be a system that users agree to participate in. Fortunately, within Enterprise Information Portals that are Portals owned by companies and used by their employees this is something easier achievable as it would be within the "free web".

Another issue is related to the interaction with expert users. There will always be a small set of expert users the system considers to be the "optimal experts". If we always represent them as experts to other users they will get overwhelmed soon. Similarly, if for certain contents only a small amount of experts is available at all, they will get overwhelmed soon, too. Incorporating load-balancing mechanisms might provide solutions to these issues.

Summary and Future Work

In this paper, we have described our approach to (implicitly) construct social networks and recommend expert users to others. We have used the combination of three metrics, namely web usage mining, tagging behavior analysis and social network analysis to construct (implicit) social networks and to determine experts. We have implemented a social network visualization component which we have incorporated into IBM's WebSphere Portal and which provides users with access to explicit contacts as well as implicit (potential) contacts, contextual contacts and expert users. The visualization component allows for a broad range of interaction with the contacts being displayed.

First evaluations with smaller deployments have shown that our system help users to find experts able to assist, and to establish networking between colleagues not knowing each other before, in order to collaborate and share knowledge significantly.

During the next few months, we intend to finalize our prototypic implementation, refine the algorithms that allow us to construct the implicit relations and use our social network models for more sophisticated recommenders and for performing content-adaptation within Portals. We additionally plan to incorporate data from other sources, such as mail- and chat servers, to enhance our social network models. Moreover, we plan to perform an in-depth evaluation of the system's performance.

References

[1] Zhang, J., Ackerman, M., 2005. Searching For Expertise in Social Networks: A Simulation of Potential Strategies. *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP'05)*, Sanibel Island, Florida, USA.

[2] McDonald, D. W., Ackerman, M.S., 2000. Expertise Recommender: A Flexible Recommendation Architecture. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW'00)*, 231-240., Philadelphia, Pennsylvania, USA.

[3] Streeter, L., Lochbaum, K., 1988. Who Knows: A System Based on Automatic Representation of Semantic Structure. *Proceedings of RIAO*, Pittsburgh, Pennsylvania, USA.

[4] Kautz, H., Selman, B., Shah, M., 1997. Referral Web: Combining social networks and collaborative filtering. *Communications of the ACM*, 40 (3). 63-65.

[5] Boey, YB., O'Madadhain, J., Smyth, P, 2004. "Netsight." http://jung.sourceforge.net/

[6] Tho, Q. T, Hui , S. C., Fong, A. C. M., 2003. A web mining approach for finding expertise in research areas. *Proceedings of International Conference on Cyberworlds (CW 2003)*, pp. 310–317, Singapore.

[7] Foner, L. Yenta, 1998. A multi-agent, referral-based matchmaking system. *Proceedings of the 1st International Conference on Autonomous Agents*, 301-307, Bologna, Italy.

[8] John, A., Seligmann, D., 2006. Collaborative Tagging and Expertise in the Enterprise. *Proceedings of Collaborative Web Tagging Workshop at WWW2006*, Edinburgh, UK.

[9] Zhang, J., Ackerman, M. S., Adamic, L., 2007. Expertise Networks in Online Communities: Structure and Algorithms. *Proceedings of the 16th International World Wide Web Conference (WWW2007)*, Banff, Canada.

[10] Liu, B., 2006. *Web Data Mining - Exploring Hyperlinks, Contents and Usage Data.* Springer.

[11] Groh, G., 2007. Groups and Group Instantiations – Detection Modeling and Applications. *Proceedings of ICSWM07*, Boulder, Colorado, USA.

[12] Groh, G., Ehmig, C., 2007. Recommendations in Taste Related Domains: Collaborative Filtering vs. Social Filtering. *Proceedings of Group07*, Sanibel Island, Florida, USA

[13] Agrawal, R., R. Sirkant, 1994. Fast Algorithms for Mining Association Rules. *Proceedings of the 20th VLDB conference,* Santiago, Chile.

[14] Srikant, R., Agrawal, R, 1996. Mining Sequential Patterns: Generalizations and Performance Improvements. *Proceedings of the 5th International Conference on Extending Database Technology*, Avignon, France.

[15] IBM Lotus Quickr
http://www.ibm.com/lotus/quickr

[16] LinkedIn
http://www.linkedin.com/

[17] Xing
http://www.xing.com/

[18] IBM WebSphere Portal
http://www.ibm.com/websphere/portal