# Using Collective Intelligence for Adaptive Navigation in Web Portals

Andreas Nauerz
IBM Research and Development
Schönaicher Straße 220
71032 Böblingen, Germany
andreas.nauerz@de.ibm.com

Stefan Pietschmann
Technische Universität Dresden
Dept. of Computer Science
01062 Dresden, Germany
stefan.pietschmann@tu-dresden.de

René Pietzsch
IBM Research and Development
Schönaicher Straße 220
71032 Böblingen, Germany
renepiet@de.ibm.com

## Abstract

*Web Portals offer users a central point of access to companywide information. Initially they focused on presenting the most valuable and widely used information providing quick and efficient information access. But the amount of accessible information has grown tremendously and finding the right information has become more complex and time consuming.*

*We utilize Web 2.0 techniques to address this issue. By incorporating tagging and rating functionality in Portals we can derive users', groups' or entire communities' interests, preferences and skills which allows for reasonable recommendations and adaptations resulting in a more user-tailored Portal. Our aim is to tailor the information presented to the set of information really relevant to a user - especially with respect to a certain context. In particular we analyze users tagging behavior to understand which resources are of higher importance to them than others and provide them with easier access to them. We additionally evaluate the semantic relations between the resources (and tags) being used in order to recommend related resources (or tags) which might be of interest to users in certain situations, too. We finally utilize our knowledge about the semantic interrelation between resources (and tags) to reorder those to better match users' needs. Thus we outline a solution in which we construct a Portal's navigation structure entirely based on the communitys' tagging behavior. Of course, as we do not only analyze each single user's tagging behavior separately, but also the tagging behavior of the entire community, we make use of collective intelligence.*

**Keywords:** Adaptive Hypermedia, Adaptive Navigation, Social Systems, Social Navigation, Group Adaptation, Recommendation, Tagging, Annotation, Adaptation

## 1 Introduction

In recent years Enterprise Information Portals (EIP) have gained importance in many companies. They represent a single point of access to personalized content, services, and applications by integrating various applications and processes into one homogeneous user interface. The fact that they are constantly growing and usually contain thousands of pages of possibly relevant information poses a serious problem and is becoming a productivity threat. EIP users need to find task- and role-specific information very quickly, but they face information overload and often feel "lost in hyperspace". In particular the huge amount of content results in complex hierarchical structures (e.g. navigation structures) designed to satisfy the majority of users. However, those super-imposed structures are not necessarily compliant to the users' mental models and therefore result in long navigation paths and significant effort to find the information needed.

Due to these reasons the next generation of Portals need to behave more adaptive. Instead of providing all possible relevant information, only those relevant in a user's current context should be presented.

The recent popularity of collaboration techniques on the Internet, particularly tagging and rating, provides new means for both semantically describing Portal content as well as for reasoning about users' interests, preferences and contexts. Beside the obvious, widely-agreed-upon use of tagging, e.g. to improve search, personal organization, recommendation and spam detection [9], it can add valuable meta information and even lightweight semantics to web resources.

This work focuses on the exploitation of the *collaborative tagging* pattern for the adaptation of Portals. Thereby, we denote "tagging" as the association of words or phrases with a Portal resource (uniquely identifiable fragments, such as pages, portlets, users, emails, wiki or blog posts, etc.). Analysis of the tagging behavior allows to model interests and preferences of users as well as semantic relations between resources, and thus to perform reasonable recommendations and adaptations. In particular, resources of higher importance to users (with respect to a certain context) can be identified and recommended or provided easier access to. We finally utilize our knowledge about the semantic distance between resources to reorder them to minimize navigation paths. By taking into account not only each single user's tagging behavior but the entire community, our recommendation and adaptation techniques benefit from the collective intelligence of all Portal users.

In the following sections we first give an overview of relevant research done in the field of adaptive Portals and tagging-based adaptation. We then present our concepts for modeling users' (or entire communitys') interests by analyzing their tagging behavior and how to leverage this information to issue recommendations and to perform adaptations. Here, we focus on the assembly of a Portal's navigation structure entirely based on the communitys' tagging behavior. We conclude with a summary and outline possible future directions of our work.

## 2 Related Work

A lot of research has been done in the field of adaptive hypermedia [1], systems that build and apply user and usage models to adapt web sites to the user's context (interests, preferences, needs, goals, etc.). One possible approach to derive those models and enable adaptation is to analyze user access or interaction data, as proposed in [14] and [5]. Projects in this context include WebWatcher [7], PageGather [15] and AMACONT [4]. Especially with respect to navigation adaptation, [16] describes an approach to speed up navigation in mobile Portals significantly. Regarding latter we will show how navigation structures can be adapted by analyzing entire communitys' tagging behavior, i.e. by leveraging their collective intelligence. We regard tagging behavior analysis as a promising additional metric to perform adaptations in addition to the web mining approaches already presented in [12].

Collaborative ranking, i.e. ranking which takes into consideration entire communitys' interests, has recently become more important. Access patterns are used to assess the importance of single web pages [2]. Improved versions of the original PageRank [13] and HITS [8] algorithms have been developed (cp. FolkRank [6], CollaborativeRank [10]). So far, all these algorithms have mainly been used to improve the ranking of search results returned by search engines as response to users' queries. We will use the ideas underlying collaborative ranking to calculate recommendations and even to dynamically adapt Portal structures to better suit single users' or entire communitys' needs.

Other work focuses on personal recommendation of content based on its relatedness to certain tag terms. [18] propose a modified version of the HITS algorithm to determine experts and high-quality documents related to a given tag. Tagging systems allow not only recommending content, but also users knowledgeable in certain areas. Based on metrics like ExpertRank [3], these users could be recommended and searched. In contrast to the HITS based approach we utilize an improved metric to determine related resources.

## 3 Concepts

After a short description of the overall system architecture we present a novel approach for leveraging information gained from the analysis of users' tagging behavior to perform recommendations and adaptations in Web Portals. We explain how to extract enriched information about the available resources, how to model single users' as well as the entire communitys' interests and preferences and, finally, how this information is used to perform reasonable recommendations and adaptations. Regarding latter, we focus on the presentation of a solution that allows for the construction of a Portals' navigation model entirely based on the communitys' tagging behavior.

### 3.1 Architecture

Figure 1 provides an overview of the conceptual architecture. Portals are comprised of resources such as pages, portlets, and users. The *Annotation Layer* allows for annotating these resources, either by users or any other, e.g. programmatic, annotator. *User Models* represent users' interests and preferences inferred via web usage mining (i.e. by analyzing users' interactions with the system and the extraction of recognizable patterns), and tagging behavior analysis. We focus on the latter, as our approaches for leveraging knowledge gained via web usage mining have already been described in [12]. Similarly, *Context Models* are built, i.e. inferred from context sensor data. Utilizing user- and context models of the whole userbase facilitates community-based adaptations. The *Adaptation Layer* uses this information to adapt the Portal's base models for navigation (defining the arrangement of pages), layout model

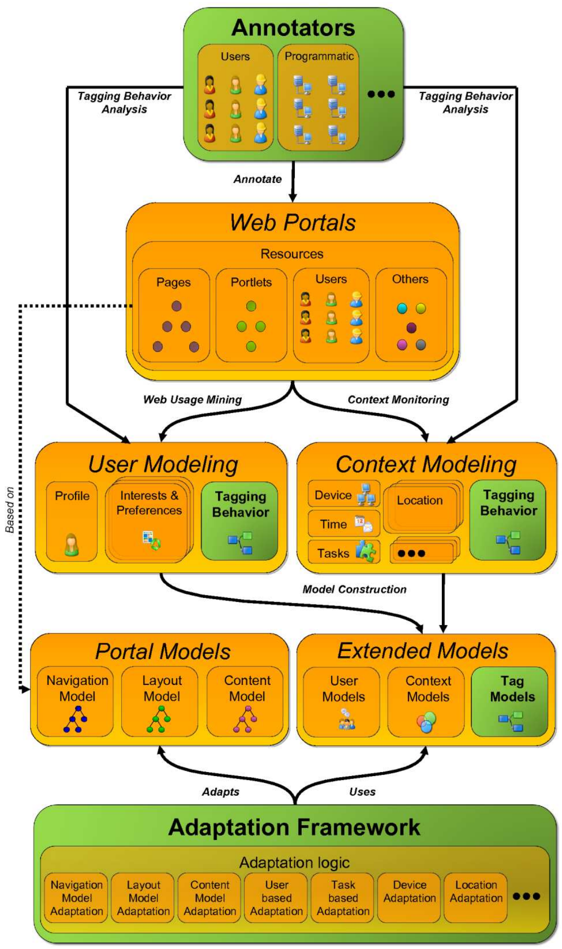


**Figure 1. Conceptual architecture**

(defining the arrangement of portlets on pages) and content. We will explain the functionality underlying, and the interplay between the layers just mentioned in the following.

## 3.2 Modeling Community's Interests

We analyze users' tagging behavior to understand both, single users' as well as the entire communitys' interests and preferences.

Tagging – the process of assigning tags to objects – has become a popular technique to describe, organize, categorize and locate resources. A tag is a (relevant) keyword or term associated with or assigned to a piece of information, thus describing the item and enabling keyword-based classification of information. Our concept allows users to annotate uniquely identifiable resources of a Portal, such as pages, portlets, and even other users. Hence, by tagging resources users can categorize content parts of the system autonomously, independent from any central instance like an administrator.

Tagging systems have proven their ability to enhance functions like search, personalization, information retrieval, and collaboration. Nearly all of these are key features in Portals. Especially with respect to searching and navigating, tagging can be regarded a promising technology. So far, navigation structures are usually created centrally by some administrator who tries to satisfy the requirements of the entire community. His decisions how to structure the system are based on his own knowledge about the users of the system, their interests and preferences, and the content being provided. Taking into consideration the size of (Enterprise Information) Portal deployments, which today often consist of 10.000s of pages used by 1.000s of users, it is unlikely that a single person can accomplish this task and estimate what a meaningful structure would be. As collective intelligence often outperforms single users' [17] we can assume that the community is able to structure content better than any administrator could.

*Tagging behavior analysis* is based on the assumption that tagging expresses interest in a resource. Hence, resources being tagged more often by a user are of higher importance to him. And since tagging is, as said, a collaborative process we can also assume that resources being tagged more often by all users are of higher importance to the entire community. Thus, analyzing users' tagging behavior allows us to better understand both, single users', as well as the entire communitys' interests and preferences.

A second assumption is that different tags being used in the system are semantically related. This means that they have a different semantic distance which can be calculated. Generally, if the **same** two tags $T_1$ and $T_2$ are applied to the **same** resources $R_1 \ldots R_n$ **often**, they **often** have a small semantic distance, or, in other words are strongly semantically related. This is obvious, as a user (or even different users) would only apply two tags to the same resource if both tags describe the information or services being offered by this resource equally well. Thus, the tags express similar semantics and are, **in most cases**, related. Understanding the semantic relation between tags we can perform various adaptations and recommendations. Regarding adaptations we can, based on tags' similarity, calculate resources' similarity and reorder resources, e.g. pages being part of the navigation, in a way such that semantically stronger related resources have a smaller click distance (cp. 3.5.1). We can further recommend related content to users based on their current selection (cp. 3.5.2). E.g. if a user has selected a page entitled *Company News* tagged with *IBM, News* we can recommend him the page *WebSphere Portal News* tagged with *IBM, News, WebSphere Portal*. Although, both pages can have a large click distance, they are, based on the applied tags, in fact semantically related.

Since tagging is a collaborative process we can, based on the semantic relation of tags, even allow for the integration of collaborative filtering-based adaptation and recommender systems that predict the utilization of a resource (page, portlet) for a particular user according to previous "ratings" by other similar users.

A third assumption is, that analyzing and comparing the tagging behavior between all users allows for partitioning them into groups of "similar behavior". Users within the same "behavioral cluster" can be provided recommendations and adaptations based on what a major subset of other users being part of the same cluster have already done. For instance, if some set of users $U_1 \ldots U_n$ always tag the same resources with similar tags we can assume that they behave similar and belong to the same "behavioral cluster". If next more than $\frac{n}{2}$ users of this cluster perform a typical action, e.g. add a specific portlet to a specific page, we can ask the remaining users if the system should perform this operation for them automatically.

Finally, by analyzing and comparing users' tagging behavior we can determine experts for certain (content) areas. Here, we can assume that users tagging certain resources have knowledge about how to deal with these. Tagging pages and portlets expresses knowledge about how to use the services provided by them, whereas tagging users expresses a relation to them. Moreover, tags applied to users might provide us with insights about their expertise. If user $U_1$ associates the tag *social-computing* with user $U_2$ he most certainly 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 $U_2$ being an expert in how to deal with these resources.

### 3.3 Modeling Community's Context

Taking into consideration only general interests neglects the context users are acting in. Common profiles could be regarded suitable models, only, if role, interests and preferences of users were not changing over time. In reality, however, interests, needs and goals change – even on a daily basis. In a business context a user might organize travels, e.g. booking flights, hotels and cars and do his travel expense. In a private scenario though, he might plan spare-time events, checking the cinema program, etc. Interests and preferences in both contexts are totally different and result in different resources being relevant to the user.

Our concept allows single users to have several context profiles between which either the system switches automatically, based on context attributes being observed (current date, time, device, location, etc.), or the user manually. The adaptation and recommendation layer utilizes both, the information stored in the user and context model, to perform its operations (i.e. to adapt Portal models such as the navigation model). Technically, the adaptation and recommendation layer partitions the user model into a sole partition for each context profile available in the context model. To determine the best matching profile, the system permanently observes a set of defined context attributes. Users always have the option to outvote the system's decision and to manually switch to another profile.

As only one context profile can be active at one specific point in time, whatever people do only influences the user model partition associated to the currently active profile. E.g. if the currently active profile is *business*, the navigation behavior does never influence the user model partition associated to the profile *private*.

The analysis of users' tagging behavior can even be used to evaluate users' context and to determine resources being of special interest in certain contexts.

Generally we can analyze how tags are applied in correlation to values of certain context attributes. For instance, we can analyze when (date and time) certain tags are applied. As an example, if a user applies the tag *private* only on Saturdays and Sundays we can assume that resources tagged with this tag are of special interest on these days only. Alternatively we can analyze which device is used when certain tags are applied. E.g. if a user applies the tag *traveling* only if using his PDA we can assume that resources tagged with this tag are of special interest when using this device.

Vice versa, we can analyze tags that already have been assigned to resources being used to determine and eventually switch the context. E.g. if a user starts to use resources mainly tagged *private* we might want to switch to the corresponding context profile.

### 3.4 Modeling Resources' Interrelation

As mentioned in the tagging behavior analysis sections, metrics that express the similarity among tags and resources are needed. In the following we provide details of our approach on how to compare tags and resources and outline further utilizations and implications.

Understanding the semantic interrelation between tags and resources, or in other words being able to calculate their semantic distance, forms the basis for our adaptation and recommendation approach. We make use of a tag-resource matrix, a compact representation of the complete tag and resource space which serves as a common source for tag as well as for resource comparison.

The semantic distance between two tags is based on cosine similarity calculations to produce a similarity value for two tags (or resources) $T_1$ and $T_2$. We define $A$ and $B$ to be the corresponding row vectors of the tag-resources matrix of the tags $T_1$ and $T_2$. The result in both cases is a number between 0 (perfect match) and $\pi$ (total opposite).

Thereby, cosine-similarity based calculation methods even allow to decide which relationships (between tags) are stronger than others. E. g., for three tags $T_1$, $T_2$ and $T_3$ we can get the distances between $T_1 - T_2$, $T_1 - T_3$, and $T_2 - T_3$. The calculation may reveal that some of the pairs have weaker relations than others ($dist(T_1, T_2) << dist(T_1, T_3)$). This finding allows us to draw the conclusion that $T_1 - T_2$ share more common resources than $T_1 - T_3$ do, and that $T_2$ is more related to $T_1$ - than $T_3$ is. Likewise, relations of resources can be calculated and evaluated.

Utilizing the outlined semantic distance functions we are able to regain structural information amongst tags or resources. Therefore we perform semantic distance calculations between all available tags (or resources). The result is interpreted as a weighted graph having all tags (or resources) as its vertices. Every vertex is connected to all others and weighted with the semantic distance between the corresponding tags (or resources). In the next step this graph is used as input for the algorithm of Kruskal which transforms it into a minimum spanning tree (MST). This MST is a dynamic topology, created from the structure that is hidden in the interrelations of tags and resources. It further allows us to reach each of our tags (or resources) in the tree with a minimum semantic distance from an arbitrarily chosen root node.

## 3.5 Adaptation and Recommendation

Based on the (enriched) user- and context models and the similarity metrics described, the adaptation layer performs various recommendations and adaptations.

### 3.5.1 Tag-based Adaptation

We argue that the transformation of the navigation structure or page layouts can result in a more user- or community-tailored Portal. Based on the models and calculations described in the previous sections, navigation nodes (i.e. pages) can be moved or hidden depending on their relevance to the user in a certain context. Pages not tagged at all might be of less interest and can hence be placed at worse positions. Pages annotated a lot might be of higher interest and hence placed at better positions. The same way, more important portlets can be grouped at the beginning of a page. In addition, arranging pages according to their semantic distances ensures that semantically related content has a small click distance.

Thus, an alternative navigation structure can be created entirely based on the tagging behavior of the community. In contrast to the super-imposed structure, this tag-driven structure has a minimal click distance between elements according to interests and the categorization performed by the community. This way, the resulting structure better fits the communitys' mental models. In addition to that the new navigation is even able to adapt to changes in the environment. I.e. if the tagging behavior of the entire community changes, the navigation structure will change, too.

Of course, navigation model adaptation can be based on a single user's personal tag collection as well, which results in a highly personalized navigation structure. It is highly adapted to his interests (as they are tagged) and allows access to relevant resources with a minimum of navigation effort. Additionally, all categories are named like the users tags so they are already well known which reduces the cognitive workload.

Besides tagging, the adaptation can further incorporate ratings applied by users to pages as an additional metric when constructing the tag-based navigation structure. While these kinds of adaptations might still seem arguably harsh, to be less obtrusive the recommender engine presented in the next section can be modified to provide short-cuts, so that relevant, highly-related navigation nodes can be accessed more easily without interfering with the original structure.

### 3.5.2 Tag-Based Recommendation

Besides adaptations, recommendations might be issued for tags and resources, as the similarity calculations provide values for both.

Tag similarity allows us to recommend related tags, based on the currently selected one. E.g. a system might be, among others, comprised of the tags *IBM, WebSphere, Downloads*. A user might be interested in release information regarding WebSphere Portal and clicks on the tag *IBM*. But, the tag *WebSphere* would have been the better one, as it is more specific. Thus, as the user clicks the tag *IBM*, which is rather general compared to the tag *WebSphere*, much more results might be returned, especially results not being of interest. Highlighting the tag *WebSphere*, as a related tag, might point the user to a tag he would otherwise have overseen.

As we know the similarity between all tags, we can not only highlight related tags when clicked within the tag cloud but also come up with a new kind of tag cloud that lists the tags being used alphabetically and clusters them depending on their semantic relation. E.g. within a tag cloud the tags *IBM, WebSphere* form one cluster and the tags *Sports, Soccer, Basketball* another one. Tags being part of the same cluster are displayed with a lower visual distance than tags being part of different clusters.

Resource similarity allows us to recommend related resources (based on the currently selected resource). E.g. if a user has selected a page entitled *Company News* and tagged it with the tags *IBM, News*, we can recommend the page *WebSphere Portal News* tagged with tags *IBM, News, WebSphere Portal*, even if both pages have a large distance within the navigation but are semantically related, based on the tags applied.

Finally, to identify users being part of the community with a similar tagging behavior a tag-user matrix is created (comparable to the tag-resource matrix). Each column in this matrix reflects the tagging profile of a user. Calculating the semantic distance between two columns of this matrix reveals the similarity of two users in terms of their tagging history. Our work about expert user determination and implicit social network construction based on users tagging behavior is described in more detail in [11].

## 4 Conclusion and Future Work

In this paper we have presented solutions that provide Portal users with easier and faster access to relevant information. An *Annotation Layer* allows to tag Portal resources (pages, portlets, users, etc.). In the *Modeling Layer* semantic interrelations between tags and hence resources are calculated. Knowledge about user characteristics and context is derived from users' tagging behavior. The *Adaptation Layer* provides means to issue recommendations to related resources (or tags) that might be useful in the users' current contexts, and to perform various adaptations to the Portal itself. With respect to latter we have demonstrated how a tailored navigation structure can be constructed entirely based on the entire communitys' tagging behavior. This structure is dynamically adapted with respect to the tagging behavior, resulting in a community-driven evolution of the Portal's navigation structure. This way it aims to provide users with the easiest and quickest access to relevant information based on their current interests and preferences.

Initial surveys have been very promising. Recommendations and adaptations were considered useful by the majority of participants (90 and 100%, resp.), which indicates the reasonability and usefulness of our system and the underlying concepts. We are currently planning more detailed evaluations with our prototype. Future work includes the extension of our recommendation and adaptation techniques as described in 3.5.1. We are also interested in incorporating more ideas from the field of social network analysis.



## References

[1] P. Brusilovsky. Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2):87–110, 2001.

[2] J. Caverlee, L. Liu, and W. Rouse. Link-based ranking of the web with source-centric collaboration. *colcom*, 0:11, 2006.

[3] S. Farrell and T. Lau. Fringe Contacts: People-Tagging for the Enterprise. In F. Smadja, A. Tomkins, and S. Golder, editors, *Proc. of the Collaborative Web Tagging Workshop at WWW2006*, 2006.

[4] M. Hinz. *Kontextsensitive Generierung adaptiver multimedialer Webanwendungen*. Dissertation, Technische Universität Dresden, February 2008.

[5] M. Hinz, S. Pietschmann, and Z. Fiala. A framework for context modeling in adaptive web applications. *IADIS International Journal of WWW/Internet*, 5(1), June 2007.

[6] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Folkrank : A ranking algorithm for folksonomies. In *LWA*, pages 111–114, 2006.

[7] T. Joachims, D. Freitag, and T. M. Mitchell. Web Watcher: A Tour Guide for the World Wide Web. In *Proc. of the 15th Intl. Joint Conf. on Artificial Intelligence*, pages 770–777, 1997.

[8] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. In *SODA*, pages 668–677, 1998.

[9] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In *Collaborative Web Tagging Workshop at WWW2006*, May 2006.

[10] A. Michail. CollaborativeRank: Motivating People to Give Helpful and Timely Ranking Suggestions. `http://collabrank.web.cse.unsw.edu.au/collabrank.pdf`, April 2005. (work in progress).

[11] A. Nauerz and G. Groh. Implicit Social Network Constructuon and Expert User Determination in Web Portals. In *Proc. of the AAAI Spring Symposium on Social Information*, Stanford Univ., California, USA, March 2008.

[12] A. Nauerz and M. Welsch. (Context)Adaptive Navigation in Web Portals. In *Proc. of the Intl. IADIS WWW/Internet Conference 2007*, Vila Real, Portugal, October 2007.

[13] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[14] M. Perkowitz and O. Etzioni. Adaptive Web Sites: an AI Challenge. In *Proc. of the 15th Intl. Joint Conf. on Artificial Intelligence*, pages 16–23, 1997.

[15] M. Perkowitz and O. Etzioni. Towards adaptive Web sites: conceptual framework and case study. *Artif. Intell.*, 118(1-2):245–275, 2000.

[16] B. Smyth and P. Cotter. Intelligent Navigation for Mobile Internet Portals. In *Proc. of the 18th Intl. Joint Conf. on Artificial Intelligence*, Acapulco, Mexico, 2003.

[17] J. Surowiecki. The Wisdom of Crowds. *American Journal of Physics*, 75(2):190–192, February 2007.

[18] H. Wu, M. Zubair, and K. Maly. Harvesting social knowledge from folksonomies. In *Proc. of the 17th Conf. on Hypertext and hypermedia*, pages 111–114, New York, NY, USA, 2006. ACM Press.