

# Personalized Information Access in a Wiki Using Structured Tagging<sup>\*</sup>

Anmol V. Singh, Andreas Wombacher, and Karl Aberer

School of Computer and Communication Sciences,  
École Polytechnique Fédérale de Lausanne (EPFL),  
CH-1015 Lausanne, Switzerland

(`anmol.tomar, andreas.wombacher, karl.aberer`)@epfl.ch

**Abstract.** Environmental engineers from different organizations work in interdisciplinary projects having the need of information exchange. In particular, a collaborative environment with personalized access to information is needed, which supports strongly varying information needs of the users. This scenario is not fully supported by existing systems. For example, content management systems do not support personalized access to content. Social bookmarking systems are not sufficiently specific in user groups, thus, they miss context information of the bookmarks and introduce imprecision and noise. In this paper we propose to combine a collaborative content management system (wiki) with the social bookmarking idea. To improve the reuse of bookmarks, we represent additional context information in the form of structured tags.

## 1 Introduction

The Swiss Experiment [1] is an interdisciplinary project between environmental engineers and computer scientists. The aim is to provide a cost efficient e-science infrastructure for environmental research, which supports collaboration between different environmental research groups.<sup>1</sup> To design the envisioned infrastructure, the computer scientists have to understand how environmental engineers handle information and data, and how they collaborate. Our discussions with groups of environmental engineers led us to the following observations.

Our first observation is that collaboration within an environmental research project is based on written or oral communication between different project members. In case of oral communication, we observed that the essentials of the communication are collected in written form also, in the form of minutes or reports of the meeting or workshop. In our further discussion, we will call any written communication a *document*.

---

<sup>\*</sup> The work presented in this paper was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

<sup>1</sup> We are aware of the e-science initiatives, for example, in the UK [2] and the US [3]. The focus in this paper is on a very specific aspect of organizing information, which can also be applied in other scenarios.

The second observation is that a project continues for several years, and many documents are generated during that period. We note that a document exists in several versions, representing the creation process. The environmental engineers use a naïve versioning system, where they encode the version in the file name. The exchange of documents is performed mostly by email, or rarely using file servers for large GIS files. During the creation of a document, it may be edited by several parties, as a consequence of a joint writing process. Due to the lack of a versioning system, the manual integration of parallel editing is tedious, and is prone to inconsistency. Moreover, a document can refer to other documents or parts of documents, which represent a semantic dependency.

The third observation is that users manage documents differently. Many users organize their documents according to organizational structures of their company (for example, one folder per department). Another common structure is based on the project, where the documents are organized according to a specific work package (WP) structure (e.g. requirements WP, design WP). Others organize their documents according to time, for example, having one folder per year. Users may also use a combination of these approaches. A conclusion of this observation is that the infrastructure must allow a user to organize the documents in her specific way.

The fourth observation is the existence of user groups with similar information needs, within the community. We are convinced by observation, that the different users within a single user group use comparable schemes to organize their content and that these users can benefit from the organization of documents done by other members of the group.

As a conclusion of the above four observations, our requirements for organizing and accessing the content are:

- versioning of documents,
- collaborative editing of documents,
- referencing of documents and parts of document, and
- personalized organization and access of documents by facilitating user groups.

From these observations we derive that there is sufficient benefit of introducing infrastructure comparable to a document management system. However, the acceptance of such a system in this user community depends on the personalized access to documents. Since each user currently manages the documents individually in personal folders, the new system must provide a better functionality in personal organization of the documents. Otherwise the users will not use the system.

As a solution to the aforementioned requirements we propose to manage personalized data access by extending a *wiki*, i.e., a collaborative content management system, by a built-in social bookmarking tool to facilitate its specialized user group. In addition to using free tagging, we propose to use *structured tags* for adding more contextual information to a tag and thus, enabling better navigation through data. The proposed approach has been implemented as a Mediawiki<sup>2</sup> extension and has been applied in the context of the Swiss Experiment project.

---

<sup>2</sup> See: <http://www.mediawiki.org>

## 2 Evaluation of Existing Solutions

There are many existing categories of infrastructures that support collaborative knowledge management, for example, groupware solutions, content management systems, and wikis. We only consider web based systems since they easily support cross-organizational collaborations. Here, we investigate the support for the requirements derived in Section 1 with the capabilities of these infrastructures. We summarize our comparison in Table 1.

### 2.1 Document and File-Sharing Systems

Document and file sharing systems [4] enable electronic handling of documents that are produced and exchanged in a group. These systems commonly support web based centralized storage, versioning and lock mechanisms, metadata extraction, and retrieval of documents in a shared workspace. Examples of such systems are Basic Support for Cooperative Work (BSCW) [5] and Groove<sup>3</sup>. BSCW<sup>4</sup> also supports posting documents with personalized annotations and rating of documents. The documents are stored in a hierarchical manner – as files under folder structures. In these systems, documents do not inherently support references to other documents. References need to be explicitly given, for example in the metadata or the document itself.

### 2.2 Groupware

Document and file-sharing functionality is also provided by groupware solutions. Groupware [6] systems support groups of people engaged in a common task providing an interface to a shared environment. An example of a web based groupware system is eGroupWare<sup>5</sup>, a free enterprise groupware software. Groupware services can include sharing of electronic calendars, project management, collective writing, e-mail handling, workflow, electronic meetings, and other activities. In the context of our requirements, the features provided by a groupware system are comparable to those of document and file sharing systems. So, groupware systems still suffer from the same limitations presented before. The additional groupware features are orthogonal to our requirements. Recent groupware solutions may also provide a wiki (see Section 2.4).

### 2.3 Content Management Systems

Document and file sharing as well as groupware systems observe content as files. Content Management Systems (CMS) [7] have content in the form of web pages and provide a solution to create and manage them. They ensure coherent content among distributed work groups by providing centralized rules and workflows.

---

<sup>3</sup> See: <http://office.microsoft.com/en-us/groove>

<sup>4</sup> See: <http://www.bscw.de/english>

<sup>5</sup> See: <http://www.egroupware.org>

An example of a CMS is WebHare<sup>6</sup>. Apart from versioning, CMS also allow referencing within documents by providing links to other documents.

Addressing our requirement of organizing content, CMS only provides a static organization of documents, in the form of web links referencing other documents. In order to have a personalized access to documents users have to set up bookmarks in their web browser. However, saving bookmarks to the browser restricts the availability of the personalized access to a single machine and browser.

## 2.4 Web 2.0

As an extension of web based content management, Web 2.0 introduces a community aspect to CMS. This instantiates new forms of sharing and editing information. We investigate social bookmarking and wikis here.

**Extending CMS with social bookmarking.** As stated earlier, CMS does not support a personalized access to documents. The use of social bookmarking addresses this issue. *Social bookmarking* systems are web based systems which function on the phenomenon of collective information retrieval from the web, and its usage in a community [8]. Examples of such systems are del.icio.us<sup>7</sup>, which supports sharing of *bookmarks* and Bibsonomy<sup>8</sup>, which supports sharing of bookmarks and publication references [9]. Bookmarks are resources on the web which a user can save on such systems. The user annotates the bookmarks with arbitrary terms called *tags*. Tags are used to cluster bookmarks and therefore provide a user specific access. Tags are available to all users, and enable the users to discover the bookmarks of others. The systems support ranking of tags by popularity.

The combination of a CMS with a social bookmarking system sums up the benefits of content management and community sharing of an open user group. In the scenario investigated in this paper, however, we have an open but a special interest group of environmental engineers, which cannot be supported by an open social bookmarking system.

**Wiki.** A second well known development in the context of Web 2.0 are wikis [10]. A *wiki* is a collaborative web site which allows its users to freely create and edit its content. A well known wiki engine is MediaWiki<sup>9</sup>. It has a simple text syntax for creating new pages and hyperlinks for referencing other pages. It supports versioning and rollback mechanisms. Unlike groupware or content management systems, wiki pages describe a closed topic, for example, the definition of a term. They link terms used in the page to respective pages that define those terms. Apart from such navigational links, a wiki editor can also specify a structure for a cluster of pages. Still, our requirement of personalized organization of documents is also not supported inherently by a wiki. The combination of a wiki and social bookmarking allows personalized access, but suffers from the limitations of being an open user group, as discussed in the previous subsection.

<sup>6</sup> See: <http://www.webhare.net>

<sup>7</sup> See: <http://del.icio.us>

<sup>8</sup> See: <http://www.bibsonomy.org>

<sup>9</sup> See: <http://www.mediawiki.org>

**Table 1.** Comparison of the existing related work with the current system

System	Versioning	Collaborative editing	Referencing documents	Personalized organization
Document and file sharing	✓	✓	✗	✓ <sup>a</sup>
Groupware	✓	✓	✗	✗
Content management system	✓	✓	✓	✗✓ <sup>b</sup>
CMS with social bookmarking	✓	✓	✓	✗✓ <sup>c</sup>
Wiki	✓	✓	✓	✗✓ <sup>b</sup>
Semantic wiki	✓	✓	✓	✗✓ <sup>b</sup>
Proposed Approach	✓	✓	✓	✓

<sup>a</sup> BSCW 4.4 supports tagging of documents.

<sup>b</sup> bookmarking not supported by the CMS; no sharing of bookmarks.

<sup>c</sup> bookmarking for specialized user groups not supported.

**Wiki with Semantic Extension.** Another possible extension to the wiki is providing semantic annotations to wiki pages [11]. The semantic annotations explicate semantics of a wiki page, supporting better search facilities [12]. However, semantic annotations are of special advantage when they are based on an ontology derived from a consensus process of the involved users. In our problem scenario, such a consensus building is difficult and too static for the fast changing demands of the users. An alternative approach could be to generate “bottom up ontologies” so called *folksonomies* which should evolve from tags used in social bookmarking [13]. Moreover, the semantic annotations are done within a wiki page, and thus are same for all the users. Thus, our problem of personalized access to documents is not solved.

### 3 The Swiss Experiment – Tagging Within a Wiki

As part of the e-science infrastructure of the Swiss Experiment project<sup>10</sup> we develop a social bookmarking extension of the wiki. Our extension supports personalized access to documents and specialized user groups.

#### 3.1 Free Tagging

We provide a tagging functionality similar to that of existing social bookmarking systems, like del.icio.us. The resources, here, wiki pages (or simply pages), are tagged with simple textual labels. Using a simple user interface, users can annotate a wiki page with a tag, while viewing it. Moreover, we embed a tag display box in every wiki page. Thus, users can conveniently explore their annotated bookmarks, as well as of other users.

<sup>10</sup> We invite the reader to access the Swiss Experiment at [www.swiss-experiment.ch](http://www.swiss-experiment.ch) with user name `monet` and password `monet`.

An important aspect of data organization is navigation and retrieval. For this purpose, we provide multiple *views* of tags and bookmarks, classified on the basis of users, tags, and pages, as summarized in Table 2.

The tag display box on each wiki page gives three views: Related tags, All tags, and My tags (see screen shot in Figure 1). *Related tags* are all the tags used to annotate a given page by all the users. As this view is provided in each page, one finds all tags related to a page in the page itself. The view *All tags* shows all the tags in the wiki, for any user and any page. *My tags* are all the tags of the logged in user, for any page she has tagged.

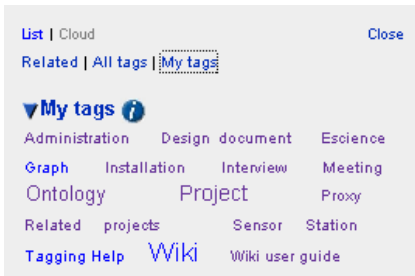
The above three views can be seen as *lists* or *tag clouds* in the tag box on each page. The tags in these views link to the respective tag pages, which is the fourth view.

Each tag has its respective page for viewing its bookmarks. A *tag page* displays all the bookmarks of any user for the particular tag. These bookmarks can be further filtered for a particular user. Thus, our fifth view displays all the bookmarks for one tag and one user.

A *user bookmarks page* lists all the wiki pages tagged by a particular user. This gives our sixth view. Moreover, one can view the bookmarks of a user restricted to one particular tag. This gives a view similar to the our fifth view.

**Table 2.** Overview of the available tag views

Views	Users	Tags	Wiki pages
Related tags	Any user	Any tag	Single page
All tags	Any user	Any tag	Any page
My tags	Single user	Any tag	Any page
Tag page	Any user	Single tag	Any page
Tag page & a user	Single user	Single tag	Any page
Bookmarks page	Single user	Any tag	Any page
Bookmarks page & a tag	Single user	Single tag	Any page



**Fig. 1.** Sample tag cloud



**Fig. 2.** Bookmarks of free tag *Interview*

### 3.2 An Example of Free Tagging

The Swiss Experiment wiki is used by the environmental engineers as well as the project members working on the Swiss Experiment infrastructure. The wiki is moderated and the content in the wiki is organized like a project web site. The wiki supports links between different wiki pages, and free text search. In addition, the editors provide a structure to access the content of the wiki. For example, the editors decided to have a category of meeting invitations and minutes, and another category of reports of interviews performed in the context of requirements elicitation. A technical user Alice, however, is not happy with this access possibility. Actually, she does not care about the invitations to the meetings, but is only interested in the minutes. Furthermore, for her, the interview reports are also minutes of meetings, and therefore should be accessible via the same category.

As discussed before, the free tagging is used to accomplish a personalized navigation for the technical user Alice. She finds the relevant pages and associates one or several tags to it. In our example, the interviews are tagged with *Interview* and *Meeting*, while the meeting minutes are tagged with *Meeting*. In Figure 2, a screen shot of a partial list of the bookmarks is depicted. In the current version of the wiki, the *Interview* tag contains 13 bookmarks and the *Meeting* tag contains about 30 bookmarks. While the 13 bookmarks are still manageable, the 30 bookmarks are much harder to look into. Additional structure would be beneficial for the navigation.

### 3.3 Structured Tagging

The Swiss experiment users deal with a lot of content which is similar in nature, but differs in values, for example, geographical coordinates and experiment numbers. We observe that free text-based tags are not sufficient to organize this content. Tagging highly parameterized content with a simple word is inadequate, especially if the number of bookmarks increases.

For better organization, we need to introduce some structure in the free text tags. The idea of grouping similar tags is already in practice, e.g., del.icio.us. But grouping does not suffice for data having experimental parameters and their values. We introduce a data structure called *complex type*. This data structure is user-defined as required. We support some pre-defined types, e.g., String, Date, Email, Geographical Coordinates, for defining a new complex type. While tagging, a tag can be associated with a defined type. The parameters of the complex type are associated with values, and the complete structure is saved for a tagged page. This adds more transparency and better organization to an otherwise flat keyword by assigning it a set of key-value pairs.

### 3.4 An Example of Structured Tagging

Extending the example in Section 3.2, the user Alice wants to provide further structure to the 30 bookmarks associated with the free tag *Meeting*. So she

date	type	location	Bookmarks
2007-12-15	interview	Davos	SwissEx:InterviewTRAMM First posted by Andreas Wombacher on 11:09, 24 July 07
2007-12-13	interview	Davos	SwissEx:SLFdatabase First posted by Andreas Wombacher on 11:09, 24 July 07
2007-06-21	meeting	Zuerich	SwissEx:Technical 2007 06 21 First posted by Andreas Wombacher on 11:35, 24 July 07
2007-06-15	interview	Davos	SwissEx:DataFiltering First posted by Andreas Wombacher on 10:58, 24 July 07
2007-06-14	meeting	skype	SwissEx:Minutes Skype Meeting 14.06.2007 First posted by Andreas Wombacher on 11:36, 24 July 07
2007-06-07	meeting	Davos	SwissEx:Minutes Davos Meeting 08.06.2007 First posted by Andreas Wombacher on 11:37, 24 July 07

Fig. 3. Bookmarks of complex tag *Meeting*

defines a complex type *Meeting Type*, which allows to associated three key-value pairs with the tag to describe the bookmark: *date* to characterize the time when the meeting take place, *type* describing whether it was a technical or a management meeting or an interview, and *location* indicating the venue of the meeting (we write *skype* for a skype phone conference).

A screen shot of the partial list of pages tagged with *Meeting* using the *Meeting Type* is depicted in Figure 3. The first three columns contain the key-value pair information and can be sorted by the user according to her needs, by clicking on the head of the table column. Adding the structure does not reduce the number of bookmarks, but provides additional sortable structure to navigate the bookmarks.

### 3.5 Architecture

We developed our solution based on the popular wiki software, MediaWiki. MediaWiki provides several hooks as extension mechanisms. *Hooks* are pre-defined points in the MediaWiki code (events) to which additional (multiple) function calls (handlers) can be assigned. We utilize the hooks for extending MediaWiki to provide the tagging features within the wiki. But our front-end requirements make the hook mechanism insufficient at certain points. For example, we need forms for user interaction for which we make additions to the front-end code of the MediaWiki.

**User interface.** We designed the user interface in such a way that the tagging module is integrated with the existing wiki front-end. This leads to a rich user experience, as one can navigate through the tags and other pages at the same time, and within the system. The development is modular, which provides the



facility to configure the tagging functionality using wiki settings. The form to tag a page is on the sidebar, so that the user can view a page while tagging it. As discussed earlier, the tags are shown in a display box on each wiki page (see Figure 1). Each tag has its own page in the wiki to show its bookmarks. Also, each user has a page for viewing her bookmarks. We created new *namespaces* to distinguish the tag pages and user bookmarks pages from other wiki pages. Namespaces are wiki's mechanism to avoid naming conflicts of pages.

**Database storage.** We develop a data model to store user tags and complex types. We do not normalize the schema and thus, introduce some redundancy to speed up queries. We store the statistics for tags, e.g., number of users for a tag and number of pages annotated by a tag, as counters in a table. This allows fast calculation of tag popularity.

Free tags are stored in the database relating the user, the bookmark, and the tag name. In case of structured tags, this representation is extended with a set of key-value pairs. In this model, a specific tag may exist as a free tag, and/or a structured tag. Moreover, a tag can be associated with multiple complex types. Thus, our storage structure supports highly flexible tagging of wiki pages.

### 3.6 Discussion

The proposed approach has been implemented and has been used in the context of the Swiss Experiment project. During the piloting, we interacted with the environmental engineers involved in the project. We concluded that user preferences, for example, how to display the tag box, are essential for usability. Furthermore, additional paradigms of navigating structured tags have to be adapted from other domains, e.g., paradigms used in data warehouses. The ability to configure the paradigm per user (or on a finer granularity) strongly enhances the usability.

As mentioned, structured tagging supports the access of bookmarks with additional information in the form of key-value pairs. However, the use of structured tags has some challenges. For example, defining a structure is an elaborate process with additional user interaction. Moreover, the flexibility to create new complex types can result in several complex types having the same or similar structure. This results in a multitude of complex types, which complicates the clustering and navigation of bookmarks. Thus, this flexibility places more responsibility on the community in their tagging behavior. Also, the reuse of existing complex types is challenging when they increase in number. Therefore, clustering, duplicate detection and migration of complex types seems to be necessary.

## 4 Conclusion and Future Work

We study a scenario with collaborating environmental engineers in the context of the Swiss Experiment project, where personalized information access within

a specialized group is needed. In this paper, we present a social bookmarking tool built within a wiki, to organize content in a personalized way. The proposed approach and implementation support structured tags to add more context information to bookmarks. This additional meta information enables multi-dimensional navigation to the bookmarks in the user interface.

In future work, we will improve the user interface by extending the navigation paradigms of bookmarks corresponding to the structured tags. Based on the evaluation of the paradigms in the project, we will analyze the collected structured tags. Moreover, we are interested in the variability of structured tags in a specialized user group, and the effect of the variability on clustering and accessing bookmarks.

## References

1. SwissEx Project: Swiss experiment (2007), <http://www.swiss-experiment.ch>
2. Research Councils UK: About the uk e-science programme (2007), <http://www.rcuk.ac.uk/escience/default.htm>
3. National Science Foundation Office of Cyberinfrastructure: Cyberinfrastructure - stimulating advances in 21st century science and engineering (2007), <http://www.nsf.gov/dir/index.jsp?org=OCI>
4. Sutton, M.: Document Management for the Enterprise: Principles, Techniques, and Applications. John Wiley & Sons, Chichester (1996)
5. Bentley, R., Horstmann, T., Sikkil, K., Trevor, J.: Supporting collaborative information sharing with the WWW: The BSCW shared workspace system, pp. 63–74 (1995)
6. Ellis, C., Gibbs, S., Rein, G.: Groupware: some issues and Experiences. *Communications of the ACM* 34(1), 38–58 (1994)
7. Boiko, B.: Content Management Bible. John Wiley & Sons, Chichester (2001)
8. Hammond, T., Hannay, T., Lund, B., Scott, J.: Social Bookmarking Tools (I). *D-Lib Magazine* (2005)
9. Hotho, A., Jschke, R., Schmitz, C., Stumme, G.: BibSonomy: A social bookmark and publication sharing system. In: *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, Aalborg University Press (2006)
10. Leuf, B., Cunningham, W.: *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, Reading (2001)
11. Krötzsch, M., Vrandečić, D., Völkel, M.: Semantic mediawiki. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L. (eds.) *ISWC 2006. LNCS*, vol. 4273, Springer, Heidelberg (2006)
12. Oren, E., Delbru, R., Möller, K., Völkel, M., Handschuh, S.: Annotation and navigation in semantic wikis. In: *Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics* (2006)
13. Gruber, T.: *Ontology of folksonomy: A mash-up of apples and oranges* (2005)