From Personal Information Management Towards Collaborative Information Management in Enterprises

Analysis, Solution Concept and Description of a Prototype*

Monika Maßun

University of Regensburg
Institute for Media, Information, and Cultural Studies (IMIK)
Universitätsstr. 31
D-93053 Regensburg
monika.massun@stud.uni-regensburg.de

Abstract
Enterprises as well as employees struggle to encounter solutions to efficiently handle increasing amounts of electronic information. Most information management systems today are based upon top-down, mainly technical approaches and impose their own structures and procedures. This often results in an even greater confusion as complexity grows and integration with existing systems is difficult. Therefore, in this work, a bottom-up conceptual approach centering on users' needs is developed. It will be shown that enabling users to better – i.e. more flexibly – organize their personal information objects[1] consequently leads to a benefit on enterprise level. The international diagnostics company Roche Diagnostics is taken as a representative showcase to analyze the challenges on the way towards efficient information handling and to design a prototype responding to these challenges.

* Published in: Osswald, Achim; Stempfhuber, Maximilian; Wolff, Christian (eds.) (2007). Open Innovation. Proc. 10th International Symposium for Information Science. Constance: UVK, 171-187.

[1] Personal information objects are e. g. emails, documents, bookmarks, contacts, calendar entries, tasks etc. needed to perform a specific task. The term will be used consequently in this paper to refer mainly to the first three.

1 Introduction

In order to analyze the information landscape and processes at Roche Diagnostics, in 2003 an external consulting company interviewed a small department of 30 employees. Users felt that most problems resulted from email overload, redundant documents in various locations and suboptimal structures in information repositories. The consultants hereupon suggested to implement a collaborative portal technology in order to unify information access[2] and decentralize content management. However, this technology addresses the core problems (cf. chapter 2.3) only partly. The overall tendency to move content from local to online web-based repositories is clearly irreversible. Still, content has to be structured and maintained somehow. As one of the core problems, namely rigid structures is not addressed, portal technologies were not asserted as a comprehensive solution. Additionally, they were seen by users as yet another complex system they would have to learn how to use[3].

Instead, the solution was sought in enabling users to better and more flexibly organize their information on an individual level, thus influencing the existing structures and applications in a bottom-up manner. Important fields of research to support this approach are Personal Information Management (PIM) and the collaborative paradigms of Web 2.0.

Chapter 2 contains a detailed analysis of the existing problems and their causes. In Chapter 3 relevant research is discussed and evaluated regarding responses offered to the analyzed problems. Lastly, in chapter 4, the solution concept and the resulting prototype will be described.

2 Analysis: Challenges in Information Management

Before starting to analyze emerging problems caused by electronic information overload in enterprises, a basic distinction should be drawn between structured and unstructured data (Table 1). Structured data is usually organized systematically in databases and can be retrieved through syntactical metadata. On the other hand, unstructured data – expressing analogous concepts – is stored rather arbitrarily in file structures or content repositories and can be retrieved through semantic meta-

---

[2] Internal document, not meant for publishing.

[3] This has been expressed in most of the interviews conducted for the doctoral thesis and is in accordance with various experts' opinions at Roche Diagnostics' business area Centralized Diagnostics R&D.

data (Gilbert and Friedman 2006). Unstructured data is estimated to constitute 80% of the information in a company (Murphy 2006, Raghavan 2002). This work focuses mainly on the management of unstructured data, especially personal information objects.

	Structured	*Unstructured*
Representation	Discrete (rows and columns)	Binary large objects: Less-defined boundaries, less-easily addressable
Storage/Persistence	DBMS or file formats	Unmanaged, file structures or content repository
Metadata Focus	Syntax (for example, location and format)	Semantics (descriptive and other markup)
Integration Tools	ETL, federation/EII and replication	Content integration and content adaptors
Standards	Structured Query Language, Open Database Connectivity and Java Database Connectivity	Java Specification Request, and Information and Content Exchange

Table 1: Distinction of structured and unstructured data. (Source: Gilbert and Friedman 2006)

2.1 Case Study: Roche Diagnostics

For this doctoral thesis, Roche Diagnostics' business area Research and Development (R&D) has been used as a representative example for the emerging problems managing unstructured data. A survey conducted among 15 employees from different departments and functions gave an impression of the users' perspective about dealing with electronic information on a daily basis[4]. Questions were asked about several data repositories as Outlook, personal and departmental file shares, the portal platform, the document management system and the intranet. It could be observed that emails, documents and intranet pages are the information objects people mostly deal with[5]. Searches for information objects are mainly performed with the search engines offered by the different data repositories. As content-based search is not offered by default search engines like Windows search, users' usually search for filenames and in subfolders. Content-based desktop search was not very familiar among the participants. A unified access to information was sought by 6 out of 15 people, especially in the upper hierarchy (project leaders, middle management), by naming folders the same way in different repositories[6]. Furthermore, there were a

4 The survey was conducted in 2005.
5 This is also the focus of PIM research.
6 This corresponds with results of a study conducted by Boardman and Sasse (2004).

couple of terms that were used repeatedly as folder names by various people independently from each other. The main problems felt by the participants were:

- Too many data repositories
- Deep folder hierarchies, forgotten files
- Imposed structures do not meet the needs
- Different views on information cannot be expressed
- Finding latest versions of forms / templates in shared repositories
- Search for synonyms is not supported
- Data redundancy
- Isolated solutions, no common platforms
- Time-consuming search for information objects

2.2 Costs of Sub-Optimal Information Handling

In order to organize and describe the problems caused by inefficient information management the notion of costs will be used in this work. It has been inspired by a discourse of Jones (2004) and further extended for this work. Jones describes the intellectual and monetary costs generated by two scenarios: keeping useless information and not keeping useful information. The emerging costs are e. g. a person's time and effort spent in keeping or finding information, the clutter of useful and useless data and the thereby increased possibility of not retrieving relevant information. Both the scenarios and the types of costs will be complemented by some more aspects for this work. Three categories of increasing costs are distinguished. Those are…

- …the user's increased time and effort to handle information,
- …the loss of information work and
- …the increasing need of storage volume.

Regarding the last point, many people assume that the costs caused by increased storage volume are not significant due to the relatively low prices of storage hardware. What adds up to those costs, however, is the whole process architecture around it, including the administration, archiving, backup and protection of growing data volumes. Other strong arguments to keep only relevant data and thus reduce storage volume are increased retrieval difficulties in vast data amounts and legal directives for archiving periods (Schlüter 2004). Thus it is here considered as a notable cost factor[7].

[7] Although it is seen as a minor factor by Jones (2004).

2.3 Causes of Increased Costs

In this work three causes for these costs have been identified: fragmentation, rigid structures and isolation. They mirror both the findings in literature and the users' perspective expressed through the survey results (cf. chapter 2.1). The relation between causes and consequences is not monocausal but can be rather imagined like a matrix (see table 2). The remedies for these identified causes will be proposed in chapter 4 after having discussed the relevant state-of-the-art technologies and concepts (chapter 3).

<table>
<tr><th>Consequence
Cause</th><th>Increased Time Effort, e. g. ↓</th><th>Loss of Information Work, e. g. ↓</th><th>Increased Storage Volume, e. g. ↓</th></tr>
<tr><td>Fragmentation</td><td>search in multiple repositories
maintenance and navigation in differing organization structures</td><td>lack of overview
establishment of folder structures from scratch in each application</td><td>storing of email attachments in email client and on file server</td></tr>
<tr><td rowspan="2">Rigid Structures</td><td rowspan="2">search for latest versions in various folders
increased organization effort
navigation in deep hierarchies</td><td>search for latest version in various folders
loss of multiple aspects of information, only one-dimensional</td><td>storing of the same information redundantly in various folders</td></tr>
<tr><td colspan="2">→ trade-off between redundancy and loss of multiple aspects of information</td></tr>
<tr><td>Isolation</td><td>the same content is structured by various users
relevant information has to be searched independently by various users</td><td>relevant information is not shared by all users equally
valuable structures are not reusable by others</td><td>storing of the same file by many users
distribution of attachments via mailing lists</td></tr>
</table>

Table 2: Causes and consequences of high costs

2.3.1 *Fragmentation*

Fragmentation describes the heterogeneity of data formats and applications. Over the last decade, enterprises have implemented numerous data repositories, platforms and applications with partly redundant functions to manage unstructured data (Murphy 2006). Among them are e. g. Enterprise Content Management Systems

(ECM), collaborative/portal platforms, Document Management Systems (DMS), Content Management Systems (CMS), workflow systems and intranets. Only recently companies have started to struggle for a decrease of complexity and the reduction of the number of their applications (Murphy 2006). One consequence of fragmentation is the user's increasing **time effort** to search and organize information which grows in proportion with the number of different storage systems and applications. Users have to manage, maintain and switch between differing structures, search mechanisms and graphical user interfaces. In order to compile a subset of logically linked information objects, generally multiple searches in various systems have to be performed (Karger and Jones 2006).

Also, the fragmentation of information among information silos (Blechar 2005) makes it harder to gain a comprehensive overview over existing and valid documents belonging to one topic. Thus needed information might be retrieved and reused only with difficulty or not at all which leads to a **loss of information**. Another consequence of the poor overview over existing information is **redundancy** as the same information is stored in several locations by several users in order to assure future retrieval.

2.3.2 Rigid Structures

Both organizational as well as personal information management systems offer ways to structure content. Usually, these structures are based on tree-like hierarchical folder structures, e. g. Microsoft SharePoint Portal Server, Outlook or file systems. They allow only one parent node (=folder) per child (=information object). In many cases, however, an information object belongs to several topics at a time. The negligence of those manifold aspects of information leads to **information loss**. As a response to the problem of inflexible folder structures, information objects are often duplicated by users in order to have them available in various contexts (=folders) and contribute thereby to increased storage need through **redundant** documents (Schlüter 2004). Eventually, if the same information object is stored in several folders it gets harder to identify the most recent version which contributes again to **increased time effort** for search.

2.3.3 Isolation

Information can be managed by administrators or by users. Usually, only information managed by administrators is accessible by everyone. A considerable share of information organization, however, takes place on an individual level or in some cases also on group level and is task-based (Boardman and Sasse 2004). Field studies have shown that people invest a lot of time and effort in thinking about how to organize their information best, especially in their file structures (Boardman and Sasse 2004, Jones et al. 2005). In most cases, these micro-organization efforts remain invisible to the rest of the employees and lead to a **loss of information**. Therefore users have to search independently from each other for information that might have been already retrieved by others. This accounts for an **increased time effort** on an organizational level. File multiplication e. g. due to distribution of information via mailing lists – which is today the most common way to publish information – is one prominent cause for **increased storage volume** (Schlüter 2004).

3 State of the Art

Among the research areas answering different aspects of the above discussed problems are e. g. Knowledge Management[8], Information Retrieval[9], Web 2.0 and Semantic Web as well as Personal Information Management (PIM). As Web 2.0, PIM and the sub-areas metadata and taxonomies in Information Retrieval have been identified as solution-relevant for this work, they will be discussed in more detail. An overview over their strengths and weaknesses can be found in Table 4 at the end of this chapter.

3.1 Web 2.0

The term Web 2.0 was coined by Tim O'Reilly and people around him in 2004[10]. They used it to describe currently emerging web technologies that focus strongly on user participation. Only two examples of Web 2.0 concepts shall be mentioned here as they are relevant for the solution. These are Wikis and Collaborative Tagging. Wikis allow a group of authors to collaborate on articles. Content creation is itera-

---

8 Knowledge Management in a broader sense offers on the technological side e. g. Content Management Systems, Document Management Systems, Collaboration / Portal software, Groupware and Workflow systems. Those are seen as further technologies that remain in the paradigm of hierarchical structuring and therefore are not seen as solutions for the described problem focus of this work.

9 Information Retrieval investigates e. g. Textmining, Visualization, Topic Maps, Automatic Classification, Taxonomies and Metadata. Except the taxonomy and metadata concepts, they have been excluded as solutions as there are already advanced technologies offering this top-down support.

10 The expression Web 2.0 is quite controversial among experts. Critics like Tim Berners-Lee say that Web 2.0 does not convey anything new but rather reuses ideas that were already part of Web 1.0 (Berners-Lee 2006). However, in this article the term will be used to refer to the typical developments that are usually cited when talking about Web 2.0, mainly participation.

tive as people keep refining and updating the articles. A famous example for a wiki is "Wikipedia"[11], an encyclopedia established by an unlimited group of volunteers. Despite the liberal collaboration practice the quality of Wikipedia articles is competitive. This was the result of a study by the "Nature" magazine published in 2005 in which experts compared the accuracy of 42 science entries from Wikipedia and the Encyclopedia Britannica (Giles 2005 cited in McAfee 2006). This suggests that some form of self-regulation regarding quality and convergence might emerge in the collaborative process.

"Collaborative Tagging describes the process by which many users add metadata in the form of keywords to shared content." (McAfee 2006) This method of collaboratively "(re)-indexing" content has recently grown popular on the Web. A known example for collaborative tagging on the Web is the bookmarking service "Delicious"[12]. Organizing content by assigning keywords is not a new idea, however, the collaborative aspect is. Usually, some kind of "authority" assigns keywords or they are derived from the document itself. With tagging any user can assign their individual and free keywords to content. This is especially useful for content that is either too much to be indexed intellectually or when there is nobody in the role of the librarian (McAfee 2006). Organizing one's content with the help of keywords first of all benefits oneself; the benefit for others is a by-product. The synergy of personal and public benefits may lead to a high motivation to use such a system.

3.2 Personal Information Management (PIM)

PIM is a research area that strives to enhance the individual's possibilities to manage their own electronic information objects. However, enterprises usually place more emphasis on general Knowledge Management and strategic information technologies and PIM is left to the individual to cope with. There are diverging theories: some research seeks the solution in better search engines, others in better structuring possibilities for information or in the combination of both.

3.2.1 Structuring and Unification

"Users need ways to unify, simplify, and consolidate information too often fragmented by location, device, and software application." For Karger and Jones (2006) information fragmentation represents the main problem when it comes to PIM. Most current PIM systems open up a new world of data formats and structures and thereby contribute to an even higher degree of data fragmentation (Jones 2004,

[11] *http://www.wikipedia.org*

[12] *http://www.del.icio.us*

Karger and Jones 2006). The primary goal is therefore to unify personal information and not add to fragmentation. Data can be unified on three levels, the **data format level**, the **display level** and the **semantic level**. Most benefits are won with the semantic level strategy (unified namespace).

Approach	*Example*	*Operations*	*Enables*
standard datatype	text	cut/copy/paste	unified search
unified presentation	window manager, Wincuts	layout, tile, show, hide	simultaneous view of information
unified namespace		reference, dereference	list below
grouping	directories, Taskmaster	group, ungroup	organizing, browsing, simultaneous view
metadata	ID3 tags, file metadata, XML	annotate, query	search, organize, browsing
cross-reference	Web, OLE	link, traverse, embed	simultaneous view, orienteering
relations	RDF, Haystack	record named relationships	unified search, browsing, orienteering

Table 3: Approaches to unification. Source: Karger and Jones 2006.

Karger and Jones (2006) come up with two solution approaches. The first one is the **Universal Labeler (UL)** that allows users to assign a common structure to documents in various applications. It is basically a project plan linked to the original sources. Thereby it combines information management with task or project management. The effort put into project planning by the user can thus be reused directly as information organization structure. The second approach is the **Haystack Project** which offers common nomenclature and graphical user interface to organize various document types independently from the application in which they had been created. It simulates thus a web-like navigation paradigm.

3.2.2 Combination of Search and Structuring

The research prototype "Phlat"[13] is an extension for Windows Desktop Search (Cutrell et al. 2006). It merges desktop search and browsing by enriching information with associative and contextual clues like milestones in time, persons, etc. As a new feature, it offers personal tagging across storage systems.

3.3 Metadata Management

One way to achieve data unification through a unified namespace is metadata. Metadata can be considered as an approach supporting multiple classifications. The

[13] *http://research.microsoft.com/adapt/phlat/default.aspx*

Gartner Research Group investigated the significance of metadata for companies. In one of their studies (Blechar 2005) they come to the conclusion that companies need to focus on consistency, reusability and accessibility of metadata in order to improve efficiency regarding time-to-market of new products. Organizational silos, data redundancy and inconsistencies across those silos are the big challenges. A way to overcome those silos is the consistent use of metadata. However, metadata is currently stored in multiple formats and locations. Despite these findings only a minority of the businesses have implemented a metadata management solution, according to Blechar (2005).

Most data integration and federation tools focus on structured data (Gilbert and Friedman 2006). However, there are some developments towards integrating unstructured data as well. IBM and Oracle are working on systems that put a virtual layer on top of heterogeneous repositories and to provide a control point for metadata. Their goal is to deliver a consistent view on information (Murphy 2006). One group of tools are "ETL" tools (Extract, Transform, Load) which support acquisition and integration of multiple data sources and then deliver the data to one or more target databases. One example for this technology is IBM WebSphere Data Stage. Another product group is called "Virtual Data Federation" or "Enterprise Information Integration (EII)" and enables users to integrate data from multiple sources into a single virtual view while the data remains at the source. An example is IBM WebSphere Information Integrator. The newly emerging tools put stronger emphasis on understanding semantics and discovering metadata as well as inference mechanisms. Examples for vendors are Unicorn and Metatomix (Gilbert and Friedman 2006).

3.4 Conclusion

Each of these approaches addresses only a subset of the discussed problem causes (fragmentation, rigid structures, isolation; cf. chapter 2.3). The discussed applications are evaluated as follows:

Concept	*Example Application*	*Fragmentation*	*Rigid Structures*	*Isolation*
Wikis	Wikipedia		(+)	+
Collaborative Tagging	Delicious		+	+
Namespace Unification	Haystack / UL	+	(+)	
Desktop Search + Tags	Phlat	+	+	
Metadata Management	IBM Websphere Data Stage	+	+	

Table 4: Evaluation of state-of-the-art approaches

4 Increasing Efficiency of Information Work

As the discussed systems do not meet all identified needs at a time, a new approach has been developed. In this chapter it will be described how the proposed desktop application prototype "TagIt" addresses three identified causes (cf. chapter 2.3). Firstly, the access to and handling of fragmented information objects should be **unified** to enable a consolidated, cross-tool view on personal information in the sense of „one-stop-shopping". Secondly, **multiple classification** should be possible in order to allow graph-like, flexible and extensible structures. Thirdly, a **centralized, participative approach** should be part of the concept. Instead of keeping the information organization efforts within the reach of a few, it should be possible for everybody to reuse this valuable intellectual work in some way.

In contrast to centralized top-down approaches of Knowledge Management, here the user is in focus. Users are provided in the first place with a tool to more efficiently manage their personal information. As a by-product, this knowledge management activity is available for the rest of the company[14].

4.1 TagIt: From Idea to Implementation

The proposed concept combines several aspects of the discussed state-of-the-art applications. It incorporates cross-tool data unification on a semantic and display level, multiple categorizations through participative keyword assignment and a centralized, reusable repository of metadata. This metadata is constituted by standardized (=terms from existing agreed corporate taxonomies chosen by the user from pick lists) and user-generated (=ad-hoc freely created terms) tags which is considered a very new and innovative characteristic of TagIt. We chose to design a desktop instead of a web-based application because of its more comfortable use and better performance – this is meant to lower the barrier to use it. The application should be easily integrable into daily work processes.

One objective in the project was to reuse the already available tools and resources instead of replacing them. There are already a lot of applications to manage information objects. On the other hand there are existing taxonomies. What is missing is the link between them. The TagIt application provides the possibility to link both corporate and private information objects virtually from existing data repositories

[14] As far as corporate content is concerned. Of course, private information will be only retrieved by its owner.

to tags that are either chosen freely or from corporate taxonomies[15]. It can be understood as a "semantic umbrella" (see figure 1).

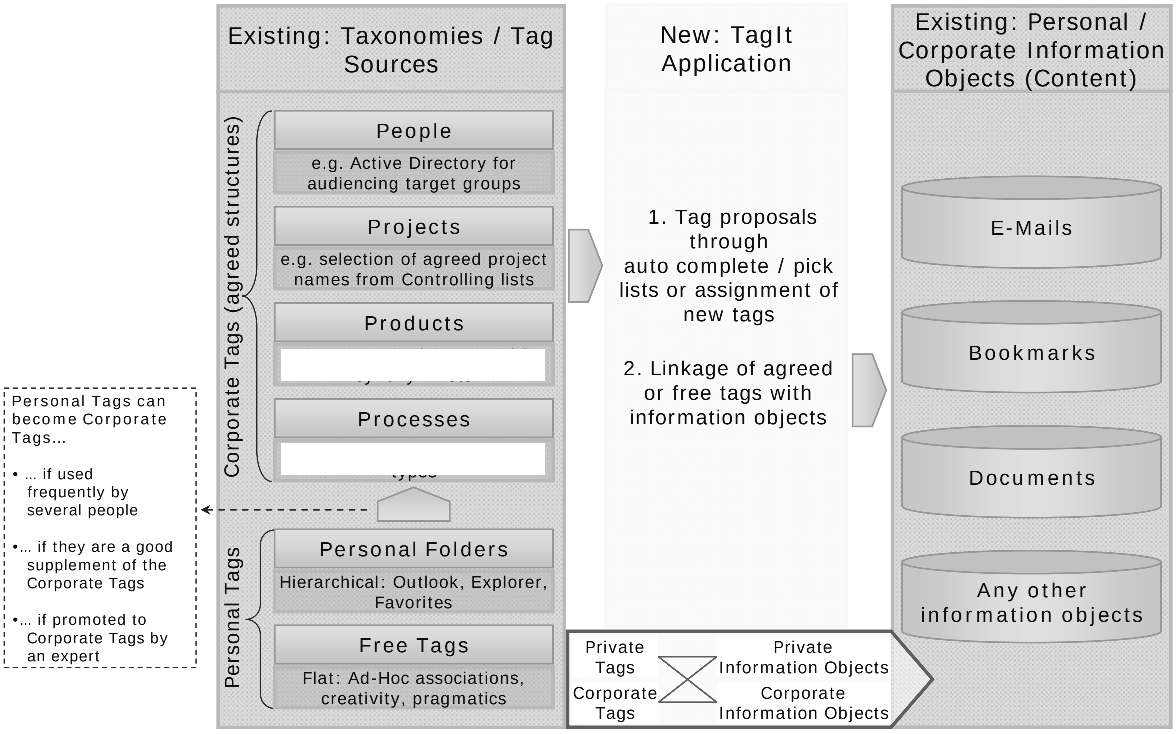

Figure 1: Existing and new components in the Roche environment

4.1.1 Cross-Tool Data Unification Instead of Fragmentation

On the main screen of the TagIt application (see figure 2) there is a three-pane view on the most frequently used information objects (email, documents and Web/intranet pages). This allows users to manage their information objects in one consistent environment instead of switching between email client, file explorer and web browser.

On the left pane there is a navigation tree including corporate and private taxonomies. The number of tagged documents is also displayed. Corporate taxonomies are imported from existing applications, e. g. product names. Private taxonomies are a user's structured tags, i.e. super- and sub-terms can be organized in a hierarchy similar to folder structures. This way of displaying tags was chosen because it is familiar to the user from Windows Explorer and Outlook. The structure can be applied to all information objects independent from their original application. By double-clicking on a tag all associated information objects are displayed in a result list (new tab).

[15] The link between information object and tag is established in a centralized database using the URIs of the information objects.

The upper pane of the screen shows the information object list, either in form of an email, file or web browser. Also the search for tags and the hit list are displayed in this area.

On the bottom pane there is a preview space for information objects. Public and private information objects can be organized in the same virtual repository and are distinguished only by their visibility to other users[16].

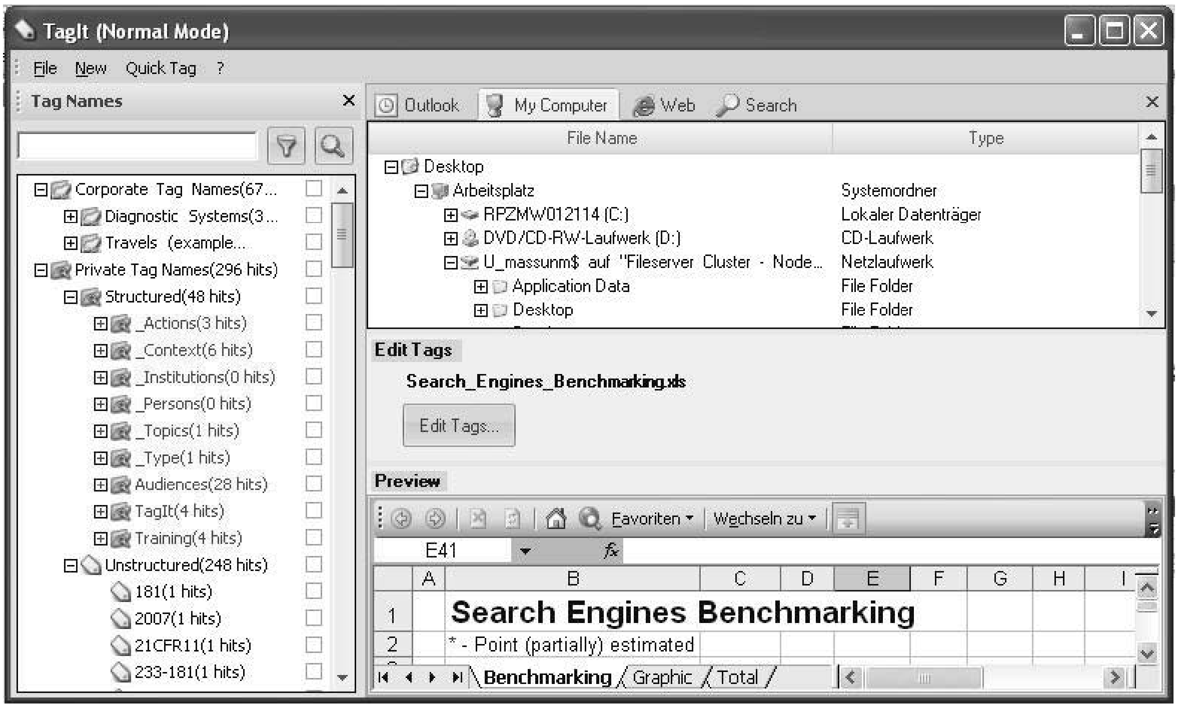

Figure 2: Three-Pane View of the TagIt application

4.1.2 Multiple Classification Instead of Rigid Structuring

As discussed in chapter 2.3.2, conventional folder structures do not allow graph-like structures in which one child can have several parent nodes (=multiple classification). Therefore in TagIt, instead of introducing yet another hierarchical organization structure, the system is extended by the possibility of **tagging**. Tagging allows users to assign multiple keywords (=tags) to information objects. They facilitate a more semantic and better structured storage of information (Lansdale 1988). Private Tags can be flat or hierarchically organized. Although the visual organization of tags can be hierarchical, the links to information objects are multi-hierarchical as one information object can be associated with many virtual folders.

[16] Private information objects like private emails or files on the personal file share are only visible to their owners. Corporate, i.e. public content is visible with all its associated tags to everyone.

Those virtual folders only offer a link to the information object but do not move it physically.

Tags can be used in a consistent way by the help of pick lists or auto-complete functionality in order to support convergence. Furthermore, an easily usable provision of corporate taxonomies is part of the prototype. Users can assign several free or agreed tags to each private or corporate information object and thus locate it in various contexts at a time. Tag assignment is possible either in a dialogue, by dragging & dropping an information object on a tag or with checkboxes (cf. figure 2).

The assignment of tag is also possible from the respective applications (Office applications, Windows Explorer, Internet Explorer) in the form of an Add-In functionality. Clicking a button in these applications invokes a TagIt dialogue and tags can be assigned to the current document.

4.1.3 Centralized User-Generated Metadata Instead of Isolation

The third conceptual component is the aspect of collective intelligence (Weiss 2005, chapter 3.1). As everybody has his or her own view on information it should be possible for all these views not only to coexist but moreover to contribute to a richer metadata environment accessible by all. This collaborative information network can produce high quality outcomes as shown in chapter 3.1. As mentioned in chapter 2.1 it can be assumed that some kind of convergence emerges as users are already using the same folder names independently from each other.

4.1.4 Search

Tag-based search is possible in various ways. One can search in private or public information objects or in both. Also, the user can choose whether to search for tags assigned by him- or herself or by other users.

Search possibilities include

- combined searches, i.e. several tags can be connected with Boolean operators,
- search with wildcards,
- search and navigate, i.e. the user can always click on a tag and change the result list by a double click or
- search with suggested terms, i.e. while typing in a search term the autocomplete functionality shows a list of all available tags starting with the typed letters from Corporate and Private Taxonomies as well as other users' tags.

As for security issues, only the information objects a certain user has access to will be retrieved. Private emails or documents of other users are certainly not retrieved.

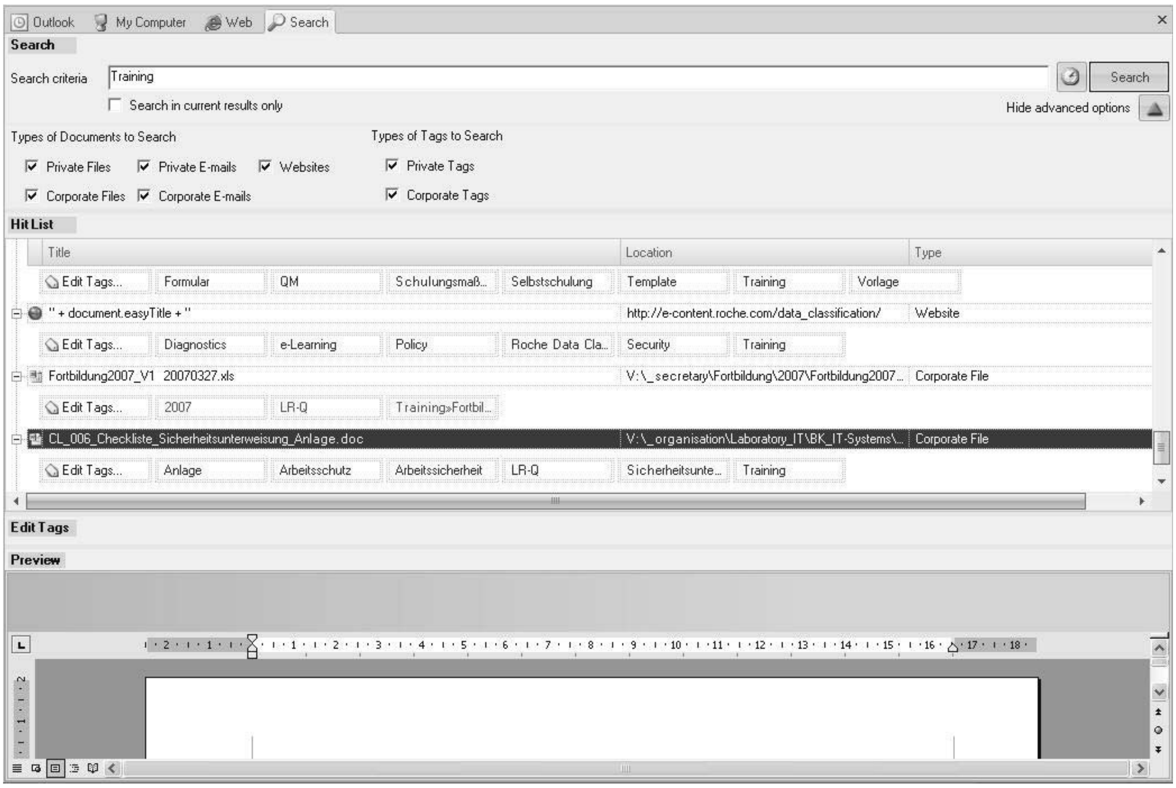

Figure 3: Search interface of TagIt

5 Prospects

The pilot will be evaluated in the first months of 2007. This evaluation will be based mainly on user studies. There will be between 10 to 15 pilot users from different areas and in different functions. The target group is project managers and information experts. Ideally, the group of pilot users can be divided into highly structured workers and "chaotic" researchers. Then, the system will be measured against their specific needs. The focus lies on the comparison between effectiveness of information retrieval tasks within the old systems and with the help of the new system. The results may be part of the presentation at ISI 2007.

After the pilot phase the user feedback will be analyzed and used to adapt the system. Apart from this evaluation there are further functionalities that are planned for programming but still have to be analyzed. One important element is the search and tagging with the help of synonym lists provided by a vocabulary server. Also, tags shall be used for ranking mechanisms and the tag-based search shall be integrated with content-based / desktop search.

6 References

Berners-Lee, T. (2006). IBM DeveloperWorks Interview with Scott Laningham. Retrieved July 28, 2006 from http://www-128.ibm.com/developerworks/podcast/dwi/cm-int082206.txt.

Blechar, M. (2005). *How to Manage Your Metadata.* Gartner Research.

Boardman, R.; Spence, R.; Sasse, M. (2004). *Stuff goes into the computer and doesn't come out: A Cross-tool study of personal information management.* In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004) (Vienna, Austria, Apr. 24-29), ACM Press, New York.

Brynko, B. (2006). *Top 10 Technology Trends.* Retrieved Sept 20, 2006 from http://pqasb.pqarchiver.com/infotoday/1001218191.html?MAC=a9cc51b746064f49c0e95f858997a4fb&did=1001218191&FMT=TG&FMTS=FT:TG:PAGE&date=Mar+2006&author=Barbara+Brynko&pub=Information+Today&printformat=&desc=Top+10+Technology+Trends.

Cormier, P. (2006). *Social Tagging For The Enterprise.* Retrieved Sept 20, 2006 from http://imbok.blogspot.com/2006/02/social-tagging-for-enterprise.html.

Cutrell, E.; Robbins, D.; Dumais, S.; Sarin, R. (2006). *Fast, Flexible Filtering with Phlat – Personal Search and Organization Made Easy.* In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004) (Montréal, Québec, Canada, Apr. 22-27), ACM Press, New York.

Dye, J. (2006). *Folksonomy: A Game of High-tech (and High-stakes) Tag.* Retrieved Sept 20, 2006 from http://pqasb.pqarchiver.com/infotoday/1012273841.html?MAC=4dea07fa3ebce9fd674a357ca2a879c2&did=1012273841&FMT=TG&FMTS=FT:TG:PAGE&date=Apr+2006&author=Jessica+Dye&pub=EContent&printformat=&desc=Folksonomy%3A+A+Game+of+High-tech+(and+High-stakes)+Tag.

Fichter, D. (2006). *Intranet Applications for Tagging and Folksonomies.*, Retrieved Sept 20, 2006 from http://pqasb.pqarchiver.com/infotoday/1033181571.html?MAC=7997db9f472d5ea9539fbf409fab7659&did=1033181571&FMT=TG&FMTS=FT:TG:PAGE&date=May%2FJun+2006&author=Darlene+Fichter&pub=Online&printformat=&desc=Intranet+Applications+for+Tagging+and+Folksonomies.

Gilbert, M.; Friedman, T. (2006). The New Data Integration Frontier: Unifying Structured and Unstructured Data. Gartner Research.

Giles J. (2005). *Internet Encyclopedias Go Head To Head.* In: Nature 438, 900-901.

Golder, S.; Huberman, B. (2006). *Usage Patterns of Collaborative Tagging Systems.* In: Journal of Information Science, 32(2), 198-208.

Hinchcliffe, D. (2006). *Enable richer business outcomes: Free your intranet with Web 2.0.*, Retrieved Sept 20, 2006 from http://blogs.zdnet.com/Hinchcliffe/?cat=28.

Jones, W. (2004). *Finders, keepers? The present and future perfect in support of personal information management.* First Monday, 9(3) (March 2004). Retrieved Jan 13, 2006 from http://firstmonday.org/issues/issue9_3/jones/index.html.

Jones, W.; Phuwanartnurak, A.; Gill, R.; Bruce, H. (2005). *Don't take my folders away! Organizing personal information to get things done.* In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2005) (Portland, OR, Apr. 2-7), ACM Press, New York.

Karger, D.; Jones W. (2006). *Data Unification in Personal Information Management.* In: Communications of the ACM, 49(1), 77-82.

Kirk, J. (2006). *Enterprises reap benefits from tagging data.* Retrieved Sept 20, 2006 from http://www.itworld.com/App/255/060524tagging/.

Lansdale, M. (1988). *The psychology of personal information management.* In: Applied Ergonomics, 19(1), 55-66.

Murphy, B. (2006). *Putting Structure Around Unstructured Data.* Interview with Howard Baldwin. In: Optimize, 51.

Raghavan, P. (2002). Information Retrieval for Enterprise Content. In: Upgrade 3(3), 5-8.

Schlüter, U. (2004). Integrationshandbuch Microsoft-Netzwerk. Galileo Press.

Weiss, A. (2005). *The Power of Collective Intelligence.* In: netWorker. Beyond file-sharing: collective intelligence, 9(3), 16-23.