

Proactive Support for the Organization of Shared Workspaces Using Activity Patterns and Content Analysis

Wolfgang Prinz, Baber Zaman
Fraunhofer FIT
D-53754 Sankt Augustin Germany
+49-2241-142730

wolfgang.prinz@fit.fraunhofer.de, zamanbabar@yahoo.com

ABSTRACT

Shared workspace systems provide virtual places for self-organized and semi-structured cooperation between local and distributed team members. These cooperation systems have been adopted by a large community over the past years and the volume of managed information is increasing rapidly. However, a problem that occurs frequently is the missing user support for the workspace organization and a lack of assistance finding the right place for storing new documents and contributions. This often results in poorly organized workspaces, making it difficult to find documents. Starting with a user survey, this paper presents a solution that assists the users in finding the right location based on an analysis of recent individual and group activities combined with a content analysis of the shared workspaces. The evaluation of the system shows evidence that the combination of the two approaches provides a useful assistance for different work situations. Furthermore some unexpected effects could be observed that makes the solution also suitable for cooperative knowledge management.

Categories and Subject Descriptors

H.4.1 [Office Automation]: Groupware; H.4.3 [Communications Applications]: Information Browsers;

General Terms

Algorithms, Design, Documentation, Human Factors.

Keywords

Shared workspaces, social computing and social navigation, awareness, office and workplace, content categorization for shared workspaces, similar documents search, semantic web for shared workspaces, noise detection for shared workspaces.

1. INTRODUCTION

In recent years shared workspace systems have become a widespread tool for the support of flexible and weakly structured cooperation in teams and communities. Typical examples for such systems are BSCW [3], Hyperwave [15], Livelink [21], Groove

[14], or MS-Sharepoint [20]. Application areas for these systems are manifold such as the coordination of lectures, intra- and inter-organisational projects, or communities.

A shared workspace normally contains different types of information such as documents, pictures, URL collections, threaded discussions, or member profiles. The content of each workspace is represented as information objects arranged in a folder hierarchy. Email notification or activity icons embedded in the user interface of the system provide awareness about the activities of the workspace members.

Since shared workspace systems do not impose a fixed structure on the workspace organization, each workspace can be organized according to the needs and requirements of the cooperating team. Most preferred structures for workspace organization are project structures (work packages, meetings) or organizational structures (departments, projects). Often structures that reflect both criteria are applied. However, the aim and intention of these structures is often not immediately visible to the users who share a workspace. Although workspace or folder descriptions can be used to describe the purpose of each workspace, users are often confused about the hierarchy, resulting in the effect that they have problems in finding the adequate folder to which they can upload a new document or where they can find the appropriate information. Although the users cooperate through a shared workspace, they often fail to develop a common understanding or common conventions [19]. They do not share a common understanding of the workspace. Instead, they have multiple views and perspectives.

This is in particular true for newcomers who have not been involved in the past creation and use of a workspace. A clear indication for this is the often observed behaviour¹: Users first browse and explore several folders or documents before they actually upload a new document. Nevertheless it happens in particular in large workspaces, that users store their documents based on their personal understanding which does not correspond to the overall group policies. Another effect is that users often create new folders within a workspace for their documents, since they cannot find the right place in the existing hierarchy. Obviously, this leads to a disorganization and fragmentation of the workspaces, which makes it difficult and time consuming for other users to find information or to reorganize the workspace.

This paper discusses the design, implementation and evaluation of a Semantic Workspace Organizer (SWO). The SWO

¹ This pattern becomes apparent by analyzing the daily activity reports that are provided by some workspaces systems. This report shows all actions of users with whom a users shares workspaces.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GROUP '05, November 6–9, 2005, Sanibel Island, Florida, USA.
Copyright 2005 ACM 1-59593-223-2/05/0011...\$5.00.

combines the activity pattern of users and groups with a textual analysis of documents inside workspaces to propose users the appropriate locations for new document upload and document search.

This approach makes use of the fact, that workspaces associate meta-information with each object (e.g. document owner, document mime type, creation date, version information, etc). In addition, user activities such as reading, editing, locking, versioning, renaming, moving existing documents, or uploading new documents are stored as events. Notifications about these events are used to increase the awareness of users about the team activities within the workspace. For our approach this information is used to propose suitable upload locations based on the recent activity pattern of the individual user or the group. In addition, we apply text-mining techniques to determine the content of documents and to build a profile of the different folders of the workspaces.

The combination of user activity patterns, meta-data and text-mining techniques offers new opportunities for document management in self-organized, shared workspaces. A system can suggest appropriate location(s) for a new document by analyzing the existing workspace structure, meta-data and activity pattern of user(s). Text mining can be combined with activity pattern analysis to determine location(s) having documents with similar content. Another possible use of the activity pattern and text mining can be the detection of wrongly placed documents and suggesting correct locations, visualizing workspaces based on activity patterns [22], and finding people based on common interests.

The expected benefit of this solution is a reduction of the time needed by the users to find the appropriate place and an improved consistency of the different workspaces and folders.

This paper is structured into the following sections: the first section describes design requirements of the system based on a user survey. The second describes the resulting design and implementation of the system including its user interface, which is followed by a section on evaluation and future enhancements.

2. Design Requirements

Design requirements for the system lie in different areas including the selection of the underlying shared workspace system and the design of the activity pattern and textual analysis based on the user preferences. We describe each in the following sections.

2.1 Identification of the Base System

BSCW (Basic Support for Cooperative Work) system is a Shared Workspace system developed at Fraunhofer FIT, that is used by more than 1 million users world-wide [3] [4]. BSCW is a web based groupware system providing features for document upload, version management, member and group administration, and awareness about the activities in workspace. A key feature of BSCW is the support of self-organization, i.e. it does not enforce a certain use pattern, policy or procedure. Although this is highly appreciated by its users, it can also lead to the effect of disorganized and fragmented workspaces, if cooperating teams use the system not disciplined enough.

The latest version of BSCW (4.2), provides a set of APIs to access and manipulate workspaces from external applications

[12]. This functionality enables the development of the Semantic Workspace Organizer (SWO) as an external application to BSCW. The large user base of BSCW further enables the evaluation of the SWO in a real work setting. For these reasons BSCW has been chosen as the underlying shared workspace system.

2.2 User Requirements Survey

We started the requirement analysis performing a user survey. Aim of this survey was the identification of the user requirements regarding a support tool for the organization of shared workspaces. The survey has been distributed to 34 experienced and frequent users of BSCW within the research department. The results indicate that many users actually require a support tool that provides an upload-guidance for locations of new document uploads. A brief summary of the results is given in Table I.

The survey results indicate further that the users apply different policies to organize their shared workspaces. This results in different structures for each workspace, similar to the observations of [25]. Some workspaces are used for educational purpose, where each folder may be used for one subject/lecture, and different sub folders for slides, papers, exams, exercises and solutions etc. Other structures represent the organizational structure of a department or the work-package and meeting structure of a project. However it becomes apparent, that the folders are used to group similar documents, hence the application of a text mining and document categorization approach can be useful to provide suggestions for a document upload.

Table 1. Survey results (34 interviewee)

Purpose	Option	Result
BSCW experience	more then 2 years	56 %
Frequency of BSCW use	Daily	60 %
Frequency of document upload	Daily	60 %
Problems in finding the location for new documents	Often Sometimes	32 % 38 %
Require assistance to suggest a location for new documents	Strongly - recommended	55 %

The survey results (table 1) indicate that 70% of all users actually have either sometimes (38%) or often (32%) problems in finding the appropriate workspace for new documents. Most of the users (55%) expect that such tool that can assist them in finding locations for new documents. Asking for the criteria that should be applied for the provision of a suggestion, 77% of the users propose suggestions for a new document upload based on content similarity (table II). Therefore we decided to include text mining and document clustering methods for the realization of the SWO.

Another method for the suggestion of upload locations is the analysis of recent user and team activities in shared workspaces. This follows the assumption [5] that the activities indicate those places in the workspace that correspond to the current activities of an individual or a team. A shared workspace system offers various

kinds of activities like document creation, reading, versioning, editing, locking, attributes changing (renaming), inviting new people in certain folders, etc. To select those activities that are most suitable for an interpretation by the SWO we have asked the users to select the activities of which they believe that they are useful in predicting the location for document uploading. Table 2 summarizes the results of this survey by showing the percentage of the results for the users who selected “very often” from different available options².

Table 2. Results of the survey to determine the appropriate factors for the calculation of the activity pattern.

Criteria (activity) to consider	Percentage (<i>Very often</i> option selected)
Content Similarity	77 %
Create activity	70 %
Version activity	60 %
Edit activity	50 %
Replace activity	46 %
Folder where last document was created	45 %
Read own created documents	14 %
Move activity	10 %
Document MIME type	10 %

Table 2 shows that most users suggest that the system should focus on create and edit operations when proposing upload locations based on the groups activity pattern. In the following section we describe the concept and implementation of the Semantic Workspace Organizer.

3. Semantic Workspace Organizer

This section describes the architecture and design of the Semantic Workspace Organizer (SWO), which is informed by the feed-back of the initial requirements survey.

3.1 System Architecture

The system architecture is illustrated in Figure 1. The SWO system is designed as a client server application using the JAVA programming language. The SWO-Server explores the shared workspaces by performing the activity pattern and content analysis to examine the suggestion results. The information is represented using Semantic Web Ontology. The SWO-client realizes the user interface accessing the server through web-services.

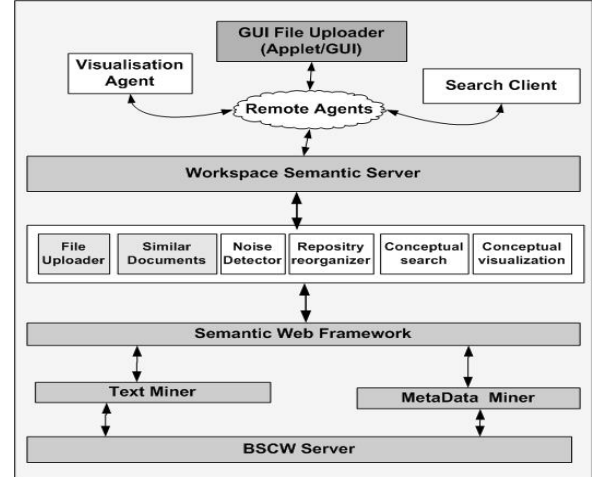


Figure 1. System architecture of the SWO

The primary components of the SWO-server are the text miner and the meta-data miner. The text miner component analyses the documents stored in the shared workspace to extract keywords and to build the index that is later used for text categorization and search. The metadata-miner component extracts the metadata and activity patterns of the user(s) from workspace. The activity pattern is represented by events stored with each document in the shared workspace. These events represent information about different types of user access (read, write, replace, create, etc) on the shared documents.

The semantic web framework component combines the results of the text-miner and metadata-miner component to build the ontology. Once the semantic web framework (ontology) is built, it can be used for a variety of purposes like suggestions for document uploading, noise detection in workspace, reorganizing the workspace using text clustering techniques, search and visualization. The workspace semantic server realizes the server side component, which is responsible for communication with remote clients.

The use of Semantic Web methods in the SWO-server increases the modularity of the application by providing a standardized interface. The SWO-Server interface can be accessed through web-services, providing the ontology of all underlying information, enabling the SWO-Clients to access and process this ontology.

3.2 Ontology

An Ontology is a formal and consensual specification of conceptualizations providing a shared and common understanding of a domain that can be communicated across people and application systems. The ontology represents the underlying data structure used in the semantic web. Two aspects of an ontology bring the semantic web to its full potential. First, the Ontology defines formal semantics of the information, enabling information processing by computers and other machines. Secondly, the Ontology can define real world semantics enabling the linking of a machine-processable content with meanings for human use, based on consensual terminology [6].

² The available options were: Very often, Often, Some times, Rarely, Never

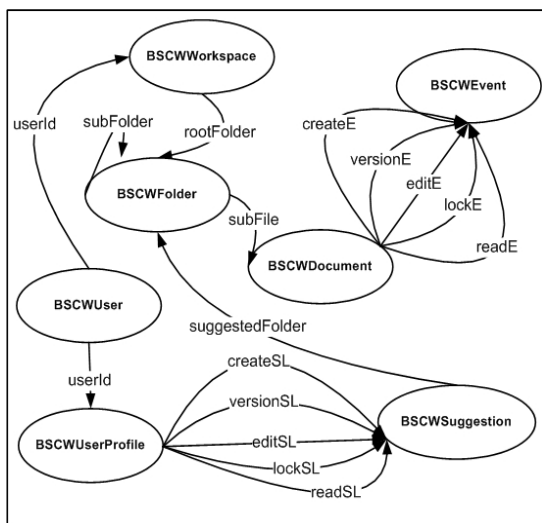


Figure 2. Ontology concepts with their relations

For our application the following concepts (figure 2) are applied: BSCWFolder, BSCWDocument, BSCWUser, BSCW-Event, BSCWWorkspace, BSCWUserProfile, BSCWSuggestion. Out of many relations and properties for each concept, we are describing only a few here which are used to represent the workspace and folder hierarchies:

```
BSCWorkspace hasRootFolder BSCWFolder
BSCWFolder hasChildFolder BSCWFolder (cardinality > 0)
BSCWFolder hasChildFile BSCWFile (cardinality > 0)
```

Each user has a Workspace, which is the default home directory, where all accessible folders appear as children folders. Users may create new folders in their workspace and access rights are transferred to children folders using inheritance mechanisms. If a user creates a new folder in his root directory, it is not visible to others, unless the folder is shared by inviting users as members to this folder.

However, shared workspace systems like BSCW or Presto [9] are not based on a canonical hierarchy of folders. Users may have different access paths to a shared folder, depending on their membership to groups and their access rights. To overcome this problem, we have designed the ontology with separate files for each folder, connected by links representing the folder hierarchy. Thus for each folder, only one instance exists, which is shared by different users and folders. Accordingly the update synchronization requires less time while the time needed for the recursive computation of the folders accessible by a user does not result in a measurable overhead and delay.

In our approach we are using the term “workspace” for the whole set of folders hierarchy starting from home folder for each user. A “workspace” becomes a “shared workspace” when a group of users can access it

and their weights are shown in Table 3. Similar approaches of activity weighting can be found in systems to support activity awareness [18],[13], [23].

All folders shared by users are examined to determine their activity rank based on the recent activities performed by the specific users as well as other users with whom the folders are shared. The weight of a folder $folder_i$ is determined by summing up the weights of each file. The weight of each file $file_j$ is determined by multiplying count $count_k$ of each event k ($k = create, version, replace, move, read$) with its weight $weight_k$.

$$\text{Weight}(file_j) = \sum_k (count_k * weight_k)$$

$$\text{Weight}(folder_i) = \sum_{j=\text{files in folder}} \text{Rank}(file_j)$$

Furthermore SWO allows users to define their default folders (bookmarks), where they upload documents most of the time and always want to see these locations in suggestions provided by the system. The system provides another non-rated upload suggestion(s), which is based on the most recent activities performed. It includes the location(s) where the most recent document was uploaded, assuming that it is probable that the user will continue working on that folder.

3.5 Text Analysis Builder

The second criteria used for providing suggestions for new documents is the computation of textual similarity between the new document and existing ones to determine the folder(s), which have close resemblance with new document.

Although the shared folders appear in a hierarchical structure in the user's workspaces, it is not wise to consider this hierarchy as a true hierarchy of concepts. The folder hierarchies are not created using a conceptual structure based on the content of the folders. The survey indicates that the folders are structured using organizational, project or other structures based on the purpose of the cooperation process for which they are used. Therefore, folders with the same or similar kind of content or not necessarily hierarchically ordered, but they may exist at different locations in the workspace. Thus the textual analysis of the folders must consider each folder as a separate class (category).

Many text categorization algorithms have been developed by data mining community [1]. However, based on surveys and evaluations performed on different text categorization methods, the following generic steps are common to most approaches.

- (1) Stop words removal
- (2) Stemming using porter stemmer
- (3) Indexing using TF x IDF [24]
- (4) Dimensionality reduction feature selection
- (5) Text classification using KNN

For our implementation we use the results of the Apache Lucene Project [26] to store the indexes of all the documents found in the shared folders. The KNN algorithm [11] has been implemented to determine the category of each folder and to compute the folders that have the highest similarity to the new document.

Normally, text categorization algorithms return only the best category based on textual similarity. However, instead of giving one result (category) only, we determine several results which are all provided to users as upload suggestion(s) for new documents. Obviously this results in suggestions of folders with have only a

weak similarity with the new document, but the user evaluation has shown that this results in very interesting side-effects.

4. The SWO User Interface

This section describes the user interface elements of the application and the provided functionality. The current implementation provides suggestions for target locations for new documents based on the criteria described in previous sections. It also provides interfaces for a content based search and to find similar documents based on content similarity. Standard functions such as workspace browsing and document views are also included.

The user interface is organized in three main sections: the workspace browser, a content panel and a history panel. The content panel offers different interfaces for document view, uploading and searching.

4.1 Browser View

The workspace browser (area 1 in Figure 4) on left side of the application displays the whole hierarchy of folders in tree form (Figure 4). When a folder is selected, the documents contained in that folder are listed in the content panel on the right hand side in a Document view (area 2 in Figure 4). Double clicking on a folder in the browser or on a document in the content panel opens it by launching the default application. If a folder is right clicked on the browser view, it opens a popup menu which has menu items to create sub folder, upload the document explicitly on selected folder and menu items to add and remove the folder into default upload locations list. Folders, which are part of the favourite upload list, are represented using different colours. The workspace browser uses different icons for folder nodes in the tree based on user access rights.

The history panel (area 3 in Figure 4) on the bottom of the application shows the history of all uploaded documents in current session. The address bar (area 4 in Figure 4) shows the URL of the current workspace. This view of the browser is mainly an alternative client to the HTML interface of BSCW.

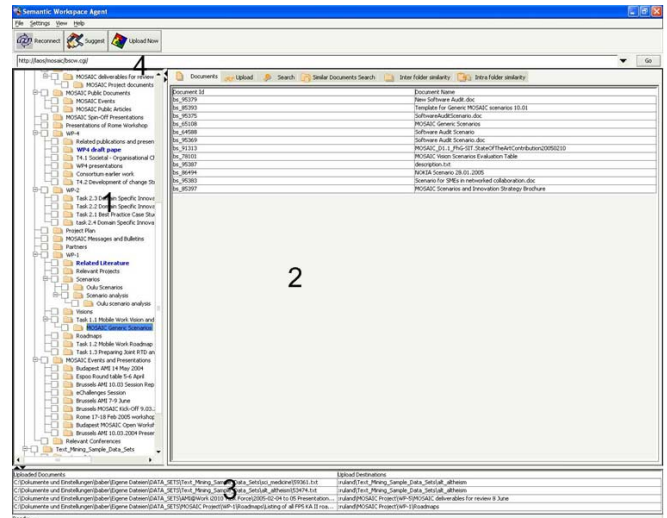


Figure 4. The browser view of the SWO client.

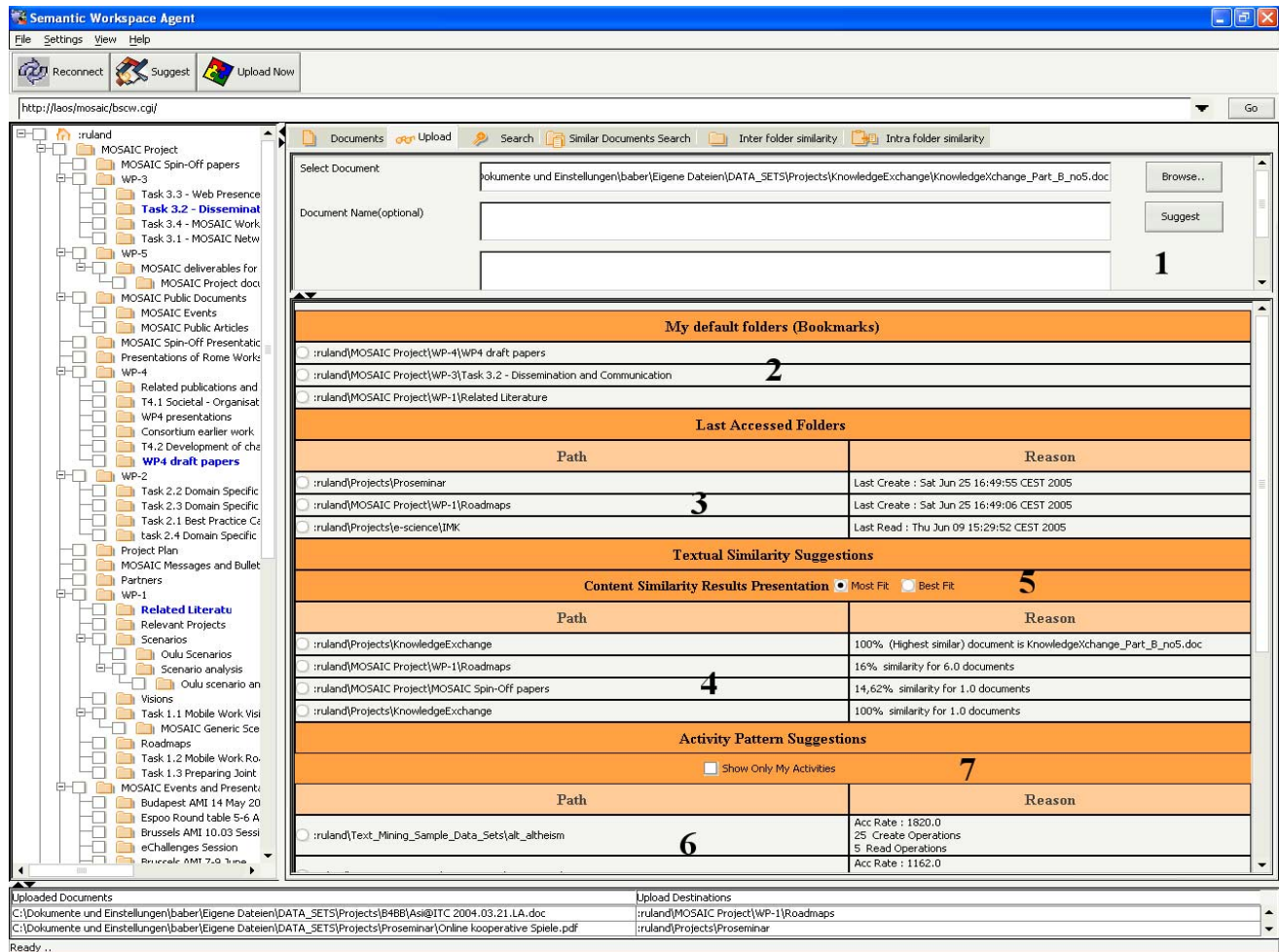


Figure 5. The upload view of the SWO client.

4.2 Upload View

The upload suggestions panel in the content panel provides an interface (area 1 in Figure 5) to browse the local file system in order to select a document for upload, to provide an alternative name for the document (optional) and a description of the document (optional). After pushing the suggest button, the document is uploaded to the SWO-server, where it is analyzed to find folders that contain similar documents. These folders are returned together with suggestions based on the user/group activity pattern as the results of the server call.

The content panel displays the results in four tables. The first table (area 2 in Figure 5) presents the upload suggestions, which user has added to its favourite list of folders. These results are presented always as upload suggestions. User can add or remove the folders into default list at any time using browser view. The second table (area 3 in figure 5) lists the folders of the user's workspaces that were most recently accessed by him or by other team members having most recent read and create activities.

The textual similarity suggestions section (area 4 in Figure 5) displays the results based on the text mining results. The results are ranked based on the maximum number of similar matches named as Most Fit. The ranking can be changed to highly similar

documents ignoring the number of results, named as Best Fit. Radio buttons (area 5 in Figure 5) can be used for this purpose.

Finally the activity pattern suggestions are listed (area 6 in Figure 5). Each row indicates the reason, why a certain location was suggested: the textual similarity results are displayed with similarity measure of the new document with suggested folder, the activity pattern suggestions display accumulative results with individual frequency of each event type in that folder. The user can switch between suggestions based only on his activities or on the activities (using checkbox area 7 in Figure 5) of all other team members in the respective workspaces.

The user can select any of these suggestions by selecting the radio button left to each suggestion, followed by pushing the upload button. If a user does not find the appropriate upload location in the list of suggestions, he can select the desired location using the workspace browser to force the application to upload the document at self selected location. This location will get ranking in next upload suggestions based on activities and highest ranking in most recent activities suggestions.

4.3 Similar Document View

This view (figure 6) supports finding similar documents in all shared workspace based on content similarity. The interface provides a button (area 1 in Figure 6) to select a document and to

displays results (area 2 in Figure 6) for similar documents. The system provides content based search which has the same view as the similar document view except that the document selection area is replaced by a text field for the query string.

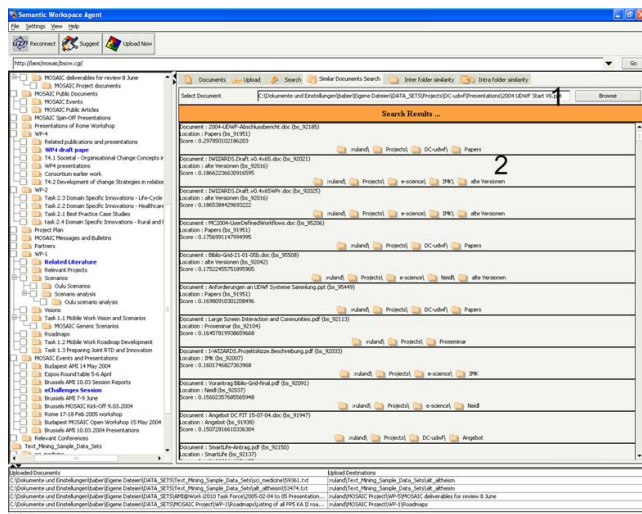


Figure 6. The similar document view of the SWO client.

4.4 Customization of the SWO

SWO has a set of defined activities and their weights (section 3.4), which are used to rate the folders while providing the suggestions for new document upload. The system provides flexibility to prioritize different activities based on personal judgment by assigning new custom weights (Figure 7). Default weights can be restored anytime by pressing the "Set Defaults" button.

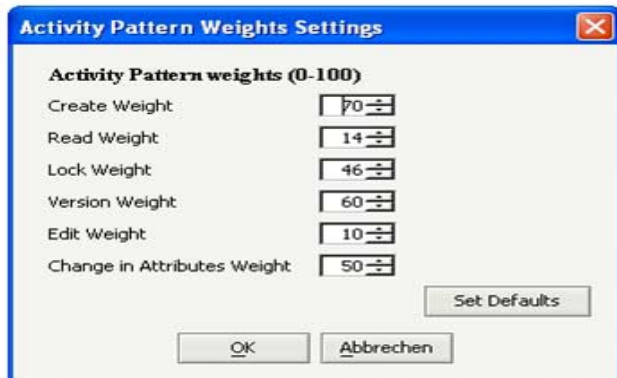


Figure 7. SWO Activity pattern weights configuration.

The choice of the suggestion criteria (favourite folders, last accessed folders, text mining, activity pattern) may be different for many users. The system allows the users to get suggestions based on the criteria they want. The upload suggestions configurations panel (Figure 8) provides an interface that can be used to include or remove the suggestions criteria while providing the suggestions for the document upload. Further the number of maximum suggestions to be provided against each criterion is also

configurable by giving the count value in the configuration dialog (Figure 8).

The configuration settings are limited to the current user only and do not effect the results of other users. These configurations are saved on the server side.



Figure 8. SWO upload suggestions criteria configuration.

5. Preliminary Evaluation of the SWO

To evaluate our approach of using background knowledge in workspaces to assist users when organizing content in shared workspaces, we have performed the evaluation in a two-phase process.

5.1 Evaluation Phase I

In the first step we have evaluated the results by distributing the application to 7 experienced BSCW users. The test scenario consisted of a set of different shared workspaces. Some of these workspaces were known to the users, i.e. they were copies of workspaces that they use every day. Other workspaces were new to the users. Therefore the users were both newcomers to a team that cooperated through workspaces already for some time as well as members of workspaces that they knew well. Beside actual project workspaces that were created and grown in the context of a real project, we also provided workspaces that contained test data for text mining algorithms. These workspaces are characterized by a high level of similarity between the documents within the workspaces and a low level of overlapping between the workspaces. Thus our test workspaces contained well-organized folders with test data, as well as "normal" project folders.

To test the application we provided a set of test documents from which the users could select documents for the 'upload' operation and the 'find similar documents' operation. The user feedback was collected by individual discussions with the test users and by a feedback form.

5.2 Evaluation Phase II

We have performed a task-oriented evaluation of the system in the second phase. The users were given a specific predefined walk through to perform and finally each user was interviewed to know the system potential, future enhancements, possible uses of the system and system usability in their work.

This phase of evaluation was performed on five different users, who also participated in the first phase. The users were given tasks to upload documents which had different level of content similarity with the existing projects in shared workspaces and differentiate the results of upload suggestions provided by the system.

Due to the limited number of users we will not provide any statistical analysis of the feedback results, but we report on qualitative findings.

5.3 Evaluation Results

All users provided a positive response to the functionality of the application and they wished to continue using the system also in the future. They do not expect that it will replace the standard BSCW user interface but would like to have the SWO functions being included there. For the future development of the SWO-client this implies that some of its functionalities should be integrated into the standard BSCW user interface, instead of adding more standard BSCW interface functionality to the SWO-client.

Except for one user, the test group said that the SWO helped them to find the right workspaces in a good or even excellent way³, although the system did not provide always the expected locations for the document upload. But even in cases where the expected folder was not suggested, the suggested location provided a good starting point from which the users could quickly browse to the final destination. This shows that the SWO can be used to support a semi-autonomous process, in which it is used to pre-select workspaces from which the user then quickly navigates to the final destination folder. This use case can save already a lot of time that is normally needed to navigate from the users root folder to the desired destination folder.

The different kind of suggestions confused some users. But after a short introduction and explanation of the interface elements, the users understood the concept and appreciated the support provided by the application.

Regarding the criteria that was used by the users to decide on the upload location the answers were more divers. Some users preferred the suggestions based on the last accessed documents, while others preferred the text similarity suggestions. This shows that the combination of both is a suitable approach to support different users' preferences as well as different cooperation situations. When users are involved in a tightly coupled cooperation process the last access suggestions provide very useful information, while this information is useless in cooperation processes where the interaction between team members occurs only between long intervals.

When evaluating the suggestions based on content similarity, two different possibilities of presenting the content similarity suggestions were discovered: A user uploaded a document in a shared workspace and afterwards the same document was uploaded again to test the system suggestions. The user expected that the location where the document has been uploaded in the first instance would be suggested as the best location. But, the results were different. The reason is that our algorithm combines the content similarity and frequency of similar documents: The textual similarity suggestions algorithm identifies the best similar documents to the given document, to suggest the location as the most appropriate which has the maximum number of the similar documents, irrespective of the similarity score.

This presentation can be named as the most-fit presentation because the ranking of the results is based on the number/frequency of similar documents out of the top best similar documents. Another presentation which is based on the value of

the similarity score irrespective of the frequency of similar documents can be named as best-fit.

The difference of the most-fit and the best-fit is shown in figure 9, where the distance of the documents in the workspace to the new document is based on the similarity score. The most-fit approach gives priority to the folder where the frequency of similar documents is greater although the average similarity is lower. The best-fit approach prioritizes the location having greater average similarity and ignoring the number of documents.

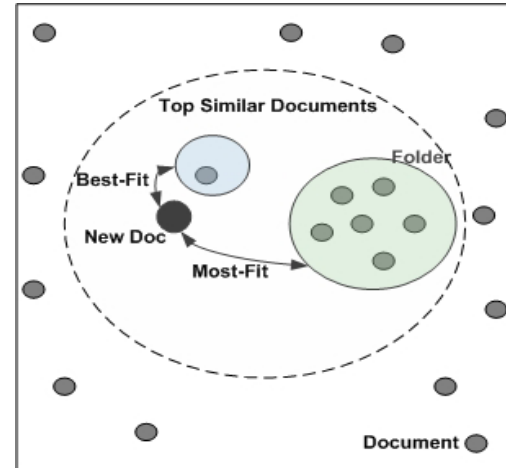


Figure 9. Best-fit and Most-fit text mining results

Accordingly, the system should be configurable to select the textual similarity behaviour. Users who are more interested in the locations where the average number of similar documents is higher can select the most-fit option, while other groups of users who prefer to get suggestions based on the most similar documents can select the best-fit option.

The suggestions based on the activity pattern analysis have not been rated as positive as the other two. We assume that this is caused by the fact that the evaluation environment did not represent the actual user's working environment. Therefore the suggestions based on this criterion were not fully understandable by the users since they were not involved in the past cooperation process. However, most users also preferred the activity pattern to be based not only on the individual behaviour but also on the activities of all team members, since this provides a better indication of the mutual cooperation process.

An interesting suggestion we got during the evaluation was to use the activity pattern based on different contexts. The context can be users-based, folders-based or a combination of both. The folders-based context should provide suggestions based on the activities performed on the set of folders, which are included in the folder-based context. Similarly, the users based context should consider activities performed by only those group of users who are defined in the user-based context. Since one user may be involved in different projects (having separate folders) and working with different people on different projects, the system should allow definition of more than one context and upload suggestions should be provided based on the selected contexts.

Some users also provided feedback on the weighting factors for the activity patterns. Here, no clear tendency can be seen. Some users prefer an almost even weighting of all activities, while others emphasize particular operations such as their individual

³ The scale for all questions was: excellent, good, normal, bad, very bad

read operations. Since the SWO enables a user specific configuration of the weighting factors, the system can be adopted to the individual preferences.

An interesting observation was made in the context of the text similarity suggestions. Some of the suggestions provided by the systems were not immediately understood by the users, resulting in the question: “Why did SWO propose that location?” However, in the same instance the users became curious about the content of that folder and they opened the workspace to have a closer look at the contained documents. Therefore a hint that was initially considered as a wrong suggestion or even a bug became a feature of the system. Users became aware of other relevant documents they have not been aware of or they may have forgotten about. Thus the SWO-system implicitly becomes also a knowledge management tool that raises the awareness of other relevant information in a shared workspace system. A further effect is, that users in some cases became aware of the disorganization and fragmentation of a workspace when different locations were proposed with a similar text analysis rating. The resulting remark was: “Well, we really should reorganize and clean this up soon ...”

6. Related Work

A lot of research on document mining and categorization has been conducted in the information retrieval community while the CSCW community has not really integrated that work in their considerations to a large scale. One exception are group-based recommendation systems [17] or the MILK system [2] that uses text mining approaches to provide related information to the documents a user is currently focusing on.

Other approaches consider the process and workflow context of a document to support a knowledge based document retrieval [8]. The work presented in this paper aims at a proactive provision of information support the organization of shared information. This concept is also used by the FXPAL Bar [7] or by Calvin [5] to provide information to users based on their current tasks and interest. However this approach takes only individual activities into account, while the SWO system is based on the analysis of cooperative activity patterns.

TaskTracer [10] and UMEA [16] are systems that also analyse user activities to contextualise resources and processes. TaskTracer applies a publish-subscribe system to monitor user interactions with a computer and to associate these events with particular user tasks. Later users can access these past activities to restore the task context. UMEA follows a similar approach by organising resources into project-related pools consisting of documents, folders, URLs, and contacts, which is similar to a BSCW workspace. The system then monitors user activities to automatically add new resources to pools associated with active projects. Both systems focus primarily on the user activities of the individual user, while our approach considers the activities of the whole group that shares a workspace. Furthermore we try to combine activity monitoring with content analysis, i.e. we aim to integrate activity related information with the activity content.

In summary we believe that the presented approach goes beyond existing approaches by integrating different methods to support a proactive organisation of shared workspaces.

7. Conclusions and Outlook

In this paper, we present a new concept of using background knowledge from shared workspaces to suggest locations for

document uploads and to support the organization of shared workspaces. The evaluation shows evidence that the approach actually supports users in the process of finding and selecting the appropriate workspace to place a new document. The user feedback shows that it is actually the combination of suggestions based on different criteria that makes the solution suitable to support different users’ preferences as well as different cooperation situations. The fact that activity patterns are based on the team behaviour is appreciated by the users, showing that the cooperative behaviour of other team members has an influence on the understanding of the shared workspaces.

As a next step we plan to integrate features of the SWO-client (suggestions, similar document retrieval) with the standard BSCW user interface, instead of adding more standard BSCW interface functionality to the SWO-client. This can be done by integrating the SWO-client functionality into the standard upload functionality of BSCW. The uploader will then return a list of suggestions from which the user can either select one or that can be used as a starting point for further navigation and exploration in the shared workspace system.

If we consider the documents produced or manipulated by a single user, instead of the documents contained in a folder, we can use our SWO server to examine categories or profiles of users. This enables a people-centred instead of a workspace-centred view on the shared workspaces. Thus users with common interests can be found by comparing user profiles with others. Similarly an expert search can be performed by providing a document and asking for all users who have produced documents similar to that one.

Shared workspaces contain lot of useful information, which is not yet used in a way, as it could be used. Currently workspaces are primarily a collection of objects in a shared folder hierarchy. This paper presents an approach to make workspaces and their organization more transparent. The analysis of activity patterns may also lead to a better understanding of the implicit and explicit policies that are applied by the team members during their cooperation process [9].

8. ACKNOWLEDGMENTS

Special thanks to all users who participated in the survey and the evaluation of the SWO system and who supported the development of the SWO system and its inter working with the BSCW system.

Further we want to thank the anonymous reviewers for their detailed and thoughtful reviews that helped us to improve the paper.

9. REFERENCES

- [1] Aas, K. and L. Eikvil, *Text categorisation: A survey*. 1999, Technical report, Norwegian Computing Center.
- [2] Agostini, A., et al. *Knowledge Management I: Stimulating knowledge discovery and sharing*. in *International ACM SIGGROUP conference on Supporting group work*. 2003. Sanibel Island, Florida, USA: ACM Press.
- [3] Appelt, W. *WWW Based Collaboration with the BSCW System*. in *SOFSEM’99*. 1999. Milovy, Czech Republic: Springer Lecture Notes in Computer Science 1725.
- [4] Appelt, W. *What Groupware Functionality do Users Really Use?* in *Proceedings of the 9th Euromicro Workshop on PDP 2001*. 2001. Mantua: IEEE Computer Society, Los Alamitos.

- [5] Bauer, T. and D.B. Leake. *Exploiting Information Access Patterns for Context-Based Retrieval*. in *International Conference on Intelligent User Interfaces, IUI-01*. 2001: ACM-Press.
- [6] Berners-Lee, T. and M. Fischetti, *Weaving the Web*. 1999, New York: Harper Collins Publishers.
- [7] Billsus, D., D.M. Hilbert, and D. Maynes-Amizade. *Improving Proactive Information Systems*. in *Intelligent User Interfaces, IUI'05*. 2005. San Diego, California, USA: ACM-Press.
- [8] Celentano, A., M.G. Fugini, and S. Pozzi, *Knowledge-based document retrieval in office environments: the Kabiria system* ACM Trans. Inf. Syst. , 1995 **13** (3): p. 237-268
- [9] Dourish, P., J. Lamping, and T. Rodden. *Building bridges: customisation and mutual intelligibility in shared category management*. in *International ACM SIGGROUP Conference on Supporting group work*. 1999. Phoenix, Arizona, USA: ACM Press.
- [10] Dragunov, A.N., et al., *TaskTracer: a desktop environment to support multi-tasking knowledge workers* in *Proceedings of the 10th international conference on Intelligent user interfaces* 2005 ACM Press: San Diego, California, USA p. 75-82
- [11] Duda, R.O. and P.E.Hart, *Pattern Classification and Scene analysis*. 1973: John Wiley & Sons.
- [12] Fraunhofer FIT and OrbiTeam, *BSCW - Basic Support for Cooperative Work*. 2005. <http://www.bscw.de>
- [13] Fuchs, L., U. Pankoke-Babatz, and W. Prinz. *Supporting Cooperative Awareness with Local Event Mechanisms: The GroupDesk System*. in *Fourth European Conference on Computer-Supported Cooperative Work: ECSCW '95*. 1995. Stockholm: Kluwer Academic Publishers.
- [14] Groove-Networks, *Groove*. 2005. <http://www.groove.net>
- [15] Hyperwave AG, *Hyperwave*. 2005. <http://www.hyperwave.com>
- [16] Kaptelinin, V., *UMEA: translating interaction histories into project contexts* in *Proceedings of the SIGCHI conference on Human factors in computing systems* 2003 ACM Press: Ft. Lauderdale, Florida, USA p. 353-360
- [17] Konstan, J., et al., *GroupLens: Applying Collaborative Filtering to Usenet News*. Communications of the ACM, 1997. **40**(3): p. 77-87.
- [18] Mariani, J. and W. Prinz, *From Multi-User to Shared Object Systems: Awareness about Co-Workers in Cooperation Support Object Databases*, in *Informatik - Wirtschaft - Gesellschaft*, H. Reichel, Editor. 1993, Springer: Berlin Heidelberg. p. 476-481.
- [19] Mark, G. and W. Prinz, *What happened to our Document in the Shared Workspace? The Need for Groupware Conventions*, in *Human-Computer Interaction INTERACT'97*, S. Howard, J. Hammond, and G. Lindgaard, Editors. 1997, Chapman&Hall: London. p. 412-420.
- [20] Microsoft, *Microsoft Sharepoint*. 2005. <http://www.sharepoint.com>
- [21] Open Text Corporation, *Livelink*. 2005. <http://www.livelink.com>
- [22] Pankoke-Babatz, U., W. Prinz, and L. Schäfer, *Stories about Asynchronous Awareness*, in *Cooperative Systems Design - Scenario-Based Design of Collaborative Systems*, F. Darses, et al., Editors. 2004, IOS Press. p. 23-38.
- [23] Rodden, T. *Populating the Application: A Model of Awareness for Cooperative Applications*. in *Conference on Computer Supported Cooperative Work (CSCW'96)*. 1996. Boston, MA.: ACM.
- [24] Salton, G. and M.J. McGill, *An Introduction to Modern Information Retrieval*. 1983: McGraw-Hill.
- [25] Simone, C. and M. Sarini. *Adaptability of Classification Schemes in Cooperation: what does it mean?* in *Seventh European Conference on Computer-Supported Cooperative Work*. 2001. Bonn, Germany: Kluwer Academic Publishers.
- [26] The Apache Jakarta Project, *Apache Lucene*. 2005.