TopicMark

A Topic-focused Bookmark Service for Professional Groups

Hui Guo and Hans-Ludwig Hausen
GMD - German National Research Center for Information Technology
FIT - Institute for Applied Information Technology
Schloss Birlinghoven
D-53754 Sankt Augustin, Germany
{hui.guo, hausen}@gmd.de

Abstract

Information overload emerges as a critical problem for information gathering as the resources on the Web grow rapidly in both scale and complexity. Networked information discovery aims to assist people's search tasks with centralized information services. User interests and user relevance judgements of Web pages with regard to these interests can be used to improve the resource discovery process. Based on this observation, an interest-topic concept model is presented. It illustrates how group processes can be enhanced with support for cooperation between users with various professional affiliation but similar interests. The design of a domain-independent bookmark service– TopicMark– is then presented that specifies the specific topic generation process, and the autonomous aggregation of information resources.

Keywords: resource discovery, topic, user interest, group process, bookmark collection, social agent.

1. Introduction & Motivation

As the Internet and the World Wide Web grow at an explosive speed, people are bothered by information overload and lacking structure when searching for interesting information sources. Information resource discovery emerges as a topic of active research with the goal of finding solutions that facilitate people's information gathering process and that meet their information needs. On the Web, various specialized collections, bookmark list pages, searchable catalogues such as Yahoo, information discovery systems such as Harvest [3] and recommender systems [13] were created to collect Web resources in one place to ease access of information. Search engines appeared as a trial to index the resources on the Web and provide retrieval services for the general public.

However, easy-to-use tools that support the users' information gathering tasks with minimal overhead are still lacking. The systems mentioned above are usually ineffective to provide relevant results with high precision for users' queries. Factors such as massive storage requirement, Web page spamming, up-to-date validation of index and ranking criteria make it difficult for search engines to keep pace with the growing Web and to provide high-quality results with regard to users' requests.

More importantly, in a professional environment, e.g. a research institute for computer science, people are accommodated to use a particular terminology (or language) when conducting information gathering tasks. This is especially true for professional groups as these groups always develop their own terminology to ease communication; the situation is similar in leisure time groups, where gangs and cliques always have their own languages. An effective personalized tool for information gathering, therefore, should allow users to use such private terminology or language. For the professional environment standardized term lists or thesauri can be used; for most hobby groups or private interest groups, their own bylaws or terms of reference can be adopted. On the other hand one might want to construct a term list or thesaurus for a particular domain or search request. Therefore termlist creation or thesaurus construction should be supported in a useful information discovery tool.

General resource discovery tools such as search engines, however, are using languages different from that of the users (or the group the user belongs to). Users don’t know how to formulate queries effectively to get better results. These tools focus on pro-

viding results with high recall but low precision, and focus on the interests of the search service provider (e.g. collect and sell information on the search engine user) without providing means to control operation or results with respect to user needs or interests. In addition, the contextual information of user search requests are not considered in the search process, thus many irrelevant results are presented to user making these services hard to use for a serious search task.

Another aspect of the information gathering process is, that people not only search information individually, but also as members of particular groups in loose or tight cooperation. To support the special information needs in a collaborative environment, an information gathering service that allows the information discovery process to be controlled by users, needs to be developed so as to support seamless integration of information systems with the resources dispersed on the Web. Wrapping search engines or digital libraries are considered ineffective to meet such needs.

This paper presents an approach to information discovery which utilizes the contextual knowledge of users interests, activities and collaboration with others. Based on this approach, web resources are harvested regarding people's shared information needs. Document indexes of the collected resources are constructed and associated with the shared user interests which we call topics. Vector space model-based retrieval can be performed within the topic-related document collections regarding the context of users' queries. The similarity among users' interests, documents' content and groups' topics are explored to support advanced information discovery processes. Such processes can be navigation in document collections by browsing topics, or recommending persons with matching interests.

There are some other ongoing research efforts adopting related technology. Subject-based search engines such as Yahoo deploy subject-focused organization of documents, but are mainly based on predefined categories or those generated from document classification. Knowledge management and search system such as Verity[9] provide means to organize intranet documents regarding user-defined categories, but lack extensibility to tackle Internet resources. Various bookmark organization tools[14] allow users to attach personal bookmarks to defined categories with little support of searching relevant documents on the Web. In the area of information retrieval, user profiles are studied and mainly used for query extension or relevance feedback. Text mining tools such as IBM[4] apply feature extraction and hierarchical clustering to sample documents so as to provide an integral view of document content without taking into account users' personal interests. Various agent-based applications [6][11] are created to support personal information discovery of Web resources by exploring user profiles.

In order to provide relevant results with both high precision and recall value, we explore the usefulness of implicitly/explicitly capturing shared interests of users in different contexts, and the construction of topic-structured collections. An Interest-Topic model is thus presented as a foundation for a "meta-search" approach to access information on the web. It is integrated in a system for "collaborative sharing and discovery " of bookmarks-TopicMark. We also examine how it can be used to augment collaboration.

Our project at the CSCW group in GMD was set up to create an open distributed platform[17] supporting various Web-situated applications using software agent technology. Under the scenario of the Social Web Programme[1], the TopicMark system built upon this platform is designed to fulfill people's various information needs during a collaboration process. It is designed to be used with low overhead, to be domain independent and self-organized, and to be an add-on service for various collaboration applications that access resources on the Web.

2. Topic Model & Group Process

This section describes the Interest-Topic model and investigates the usefulness of the model to support group processes.

2.1Users' information needs in a collaborative environment

When people work in a collaborative environment, they usually have different information needs than when working individually. In the latter case, it might be unclear what kind of information to look for and what are the proper questions to ask at the starting point. People may at first simply browse through Web pages, following interesting links and gradually come up with some specific questions or queries.

As figure 1 shows, in a collaborative context, people have very specific information needs. These needs arise from the interaction occurring during their cooperation work and closely relate to the topics they

1. URL:http://orgwis.gmd.de/projects/socialweb/

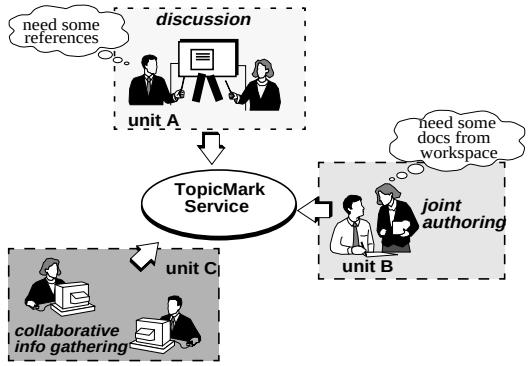

Figure 1. Information needs in collaborative environments

are focusing on. The critical requirement for such a search request is that highly relevant results are delivered in real-time so that cooperation can smoothly proceed rather than be interrupted or delayed. An information service within or across the collaborative environment should contain compact information resources relevant to people's current tasks and interests so as to meet their specific requests. In addition, people sometimes distribute the search task among each other and concentrate on particular aspects. Presenting information relevant to the overall search context will facilitate the collaborative gathering process. To this end, users' shared interests and the context can be captured to support their information gathering tasks.

2.2 Interest-Topic model

To realize the information service mentioned above, one critical problem has to be solved: organize the documents in a way that highly relevant documents are selected for a user's particular search interests. Traditional statistical classification of documents is not feasible here since it is mainly dependent on the statistical characteristics of the document collection that may be different in different users' view of the contents. Thus, search results based on this method are very likely to be irrelevant with respect to the users' real interest.

The proposed Interest-Topic model is shown in figure 2. The major concepts are:

- capturing users' personal interests in a collaborative environment.
- extraction of shared user interests as topics.
- categorization of documents to topics based on content relevance.

Within collaborative environments, users' personal interests (long-term/short-term) and special needs are captured. Internally they are formalized as text vectors. Based on user-selected terminology or community language such as ACM Classification System for expressing interests, these interests are aggregated and clustered. Shared interests are extracted as topics and normalized by a standard thesaurus.

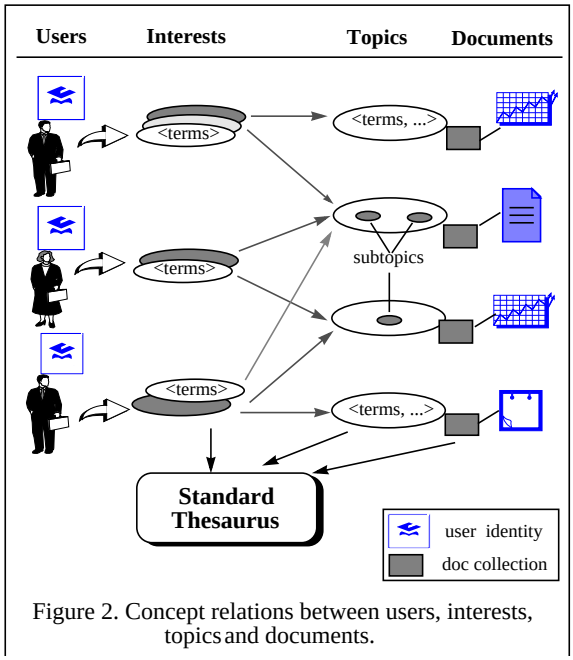

Figure 2. Concept relations between users, interests, topics and documents.

Topics are also formalized as text vectors. Regarding these generated topics, relevant web pages on the Web are harvested, collected, indexed and ranked as sample documents relevant to the topics. By analyzing the content of these web pages, important terms are extracted based on term occurrence frequency and added to topics. Incoming documents are categorized to existing topics and topic-controlled harvesting is continued. According to this approach, gradually users, user interests, topics and documents are associated to each other as illustrated in figure 2.

Documents are categorized according to the users' shared interests. When a user issues a search request, the contextual information of this request is matched to registered topics; the query itself is matched with the documents associated with those topics. Results generated in this way are expected to be relevant since they match the query based on content, and the categories that they are associated with match the context of the request. All these matching processes adopt a correlation coefficient function, e.g. cosine function or distance measure[16].

2.3 Specification of user interests

By user interest, we mean the information needs aspects of a user's work. These needs constitute the context of users' search requests. One example is the need to search some information relevant to the topics users are concerned about. To capture such needs, the three-dimension user model [15] is adopted as. Based on that, users' short-term/long-term interests both as an individual and as members of a group can be captured implicitly.

User interests can be indicated by the user's membership within various physical organizations. Firstly users' professional affiliation in the real world represents the scenario of their interests as a whole, e.g. a computer scientist's major interests fall into the computer science area. Users' institutional affiliation can serve as a symbol of interests, e.g. a researcher in GMD-FIT mainly works in the HCI or CSCW area. In addition, when people use particular collaborative applications to assist their work, the user identity in a group also reflects interests in particular contexts, e.g. a person invited to a "Social Web Program" workspace in the BSCW shared workspace system [2] should have interests in that research direction. Such characteristics of a user are the basic part of their profiles when they join a collaborative environment, and thus can be directly captured.

On the other hand, users' search interests are usually created as queries which specify the keywords or phrases to appear in the document. These queries represent a snap shot of the user interest at a given time. In the case where users are offered some means to store or organize the queries together along with the collected documents, e.g. a Netscape bookmark folder, the context of each query can be represented by the container name, depending on the way that queries and documents are grouped and the normalization to name the containers. Usually users attach several highly-rated documents as very relevant documents of particular interests. The content of these documents also behave as a description of the interests. Using query-routing technology as introduced in [10], important terms can be extracted and treated as descriptors of the user interest. Therefore a user interest consist of three element-queries, tasks, and relevant documents.

A user interest is represented as an object with some attributes: UserID, semantic context, denotation, attached documents, and timestamp. One user may have multiple interests. Each aspects of user interests are captured in the user interface and are transformed into this standard representation without losing semantic consistency. In this definition, UserID refers to the user owning this interest. Semantic context means the context in which a user delivers the search request, e.g. the task he/she is doing, the content of the document he/she is authoring. Denotation is defined as list of descriptors or terms corresponding to the user's queries. The timestamp or other attributes are used to indicate changes in the interest and thus trigger the topic-formulation process. An example in the experimental TopicMark system is:

```
Angi's interest ::=
   { ANGI, {"GMD-FIT",  "social web"},
   {{"virtual places", "avatar"},
    {"social construction of knowledge",
      "conceptual index"},
    {"collaborative filtering",
      "recommender system"}},
   list of high-rated documents, Mar.98};
```

2.4 Aggregation of interests into topics

A topic is defined as the shared aspects of a certain number of users' interests. Thus it mainly captures the collaborative aspects of the users' work. It can be identified by the frequently co-occuring terms that appear in the users' interests. Since a few community languages or public dictionaries such as FOLDOC [7] are available, only those user interests expressed in a same language are clustered to generate topics and topics are thus grouped regarding the language. Once generated, depending on whether it exists in standard thesaurus or appears in user interests as an agreed phrase, a topic is registered in existing topics group.

A topic is described in a similar way as interests:

```
Topic :: = {Users, aggregated context,
denotation, listof subtopics, attached
documents, timestamp};
```

The subtopics of a topic can be derived by looking for partially shared interests. Subtopics are also made uniform using a standard thesaurus or a user-controlled vocabulary, e.g. words appearing in users' interests as a phrase statistically. Each topic generated is associated with a collection of documents according to the relevance calculated as a correlation between the topic's description and the contents of the documents. Once a new topic is generated, the system can either attach existing documents to the topic, or initiate a documents collecting task and start content-based attaching process.

A topic can be reformulated by extracting high-frequency terms in the attached documents as additional terms in the denotation. It could also be revised once users change their interests or define new interests. A fine-grained topic structure can be defined by aggregating structured personal interests. More specifically, topics can be classified as topics for a group, or topics for several groups sharing interests.

2.5 Augmentation of collaboration

The TopicMark solution for information gathering introduces several important ways to support collaboration. Basically, since information resources on the Web are collected and organized regarding users' interests, TopicMark offers an easy-to-use central service matching users' requests accurately. Secondly, users' shared interests are captured and it is possible to classify users into groups and to bring people together or to recommend person with similar interests, e.g. recommend a technical expert. The internally identified groups can also help users cooperate across physical group boundaries. Furthermore, by exposing the generated topics to the users and by allowing them to manipulate or navigate through the topic space, users are offered a good chance of mutual learning, information discovery and constructing knowledge of their interest.

The topics generated in the TopicMark system are adapted dynamically according to the changes in users' interests so as to capture users' behavior. It is possible to build personalized agents for individual users, and socialized agents for the groups. Mediated by TopicMark system, the cooperation between various kinds of information systems on the Web can be facilitated.

To explore the usefulness of the model, we conducted some experiments using the raw data in a recommender system-LiveMarks[8]. The results shows that user interests show a lot of common aspects, thus making it feasible to construct topics. However it also exhibits the necessity to specify structure of topics. We intend to make use of all the web pages in GMD intranet indexed by TopicMark robot to achieve more understanding.

3. An information gathering service for groups – TopicMark

This section gives a detailed description of the TopicMark service based on the topic model presented before.

3.1 System functionality

TopicMark is based on vector space model[16] for information storage and retrieval. It provides:

1. Users or group-specific indexing of web resource;
2. Interest and topic-focused search and retrieval;
3. Enrichment of web resource description and automatic updation of resource’s validity.

It supports above features by registration of users, groups and web resource with interests or group topics, and doing indexing of these artifacts via user-specific term lists or thesauri.

More specifically, in TopicMark user requests and web items are registered with both interests and topics. For the indexing (i.e. the mapping of the raw web item or the request onto the set of descriptors) user-selected term lists or mature thesauri are used. In an initialization phase one can use a standard term list or thesaurus (e.g. ACM computing reviews thesaurus, ISO multilingual term list) or a dedicated subset thereof. It is assumed that indexing with user/group-approved thesauri improves user satisfaction as well as recall and precision. A request is processed by matching it against the user’s interests or topics or both or (if that match was not successful) approaching the web via public search engines and feeding the results into the interests and topics. TopicMark updates its data base on interests, topics and associated web items on defined time intervals and concurrently tests the data on these web items. This ensures availability and validity of the marked web items. Depending on user defined preferences users are notified on updates and changes on web items as well as on other users or on groups. As a consequence users participating in TopicMark will be aware of users, groups, interests, topics as well as on available and valid web items.

TopicMark can be extended by introducing other functionality. For example, the contextual information of users’ interests as well as particular web resources can be explored so as to match documents with users’ request more comprehensively and present results in a more understandable way; a group matcher can be constructed by specifying shared user interests to support group-specific information fathering. A full integration paradigm can be applied to integrate conceptual index [17], TopicMark and recommender LiveMarks.

3.2 User interface

The screendump of the TopicMark user interface is shown in figure 3. TopicMark is currently integrated with a recommender system as an add-on information service. TopicMark provides several places for users to conduct collaborative information gathering.

- Tasks page: a place to join predefined groups and express search interests in task level;
- Results page: a place to input queries, browse URLs and share annotations;
- Topics page: a place to access the captured topics;
- Search interface: a place to conduct context-related search using ACM Topic Tree .

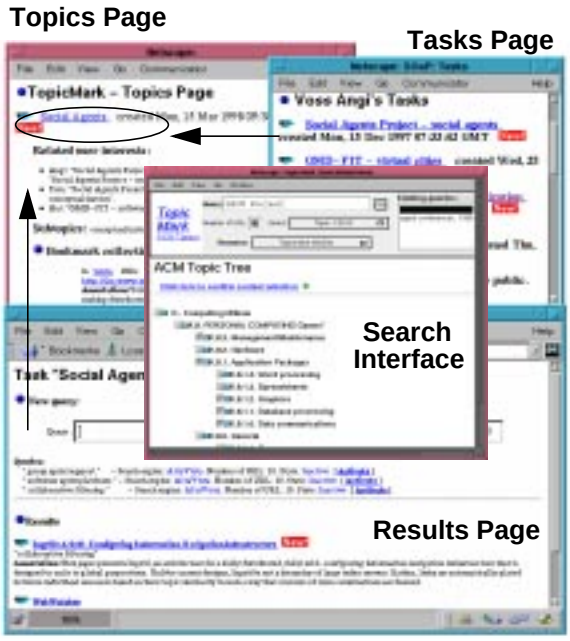

Figure 3. TopicMark User Interface

Under user-specified privacy constraints, the tasks and queries in the first two places are captured by TopicMark and are used to maintain the place for topics. As figure 4. show, the topic *Social Agents* are generated out of user Tom's personal interests and user Voss Angi's interests represented by the *"Social Agents Project - social agents"* task, queries such as *"group agent support"* and *"software agent platform"*, and the highly-rated bookmarks within results page. The topics page provides a central view for users to be aware of the activities occurring within/outside the groups by navigating through the topics and the associated information such as user interests, subtopics, and bookmark collections.

TopicMark also automatically matches users' queries against the topics and bookmark collections and puts relevant bookmarks into users' particular task page as retrieval results. In addition, in TopicMark each artifact as a part of user interests has an event indicator, e.g. when a new topic is generated regarding registered queries, the indication icon *NEW* will be highlighted in the topics place to inform the users of the event. Such events can also be "pushed" to users in order to support real-time service.

3.3 Topic, shared interest and user interest

We intend to develop TopicMark as an information gathering system for specialized groups. The system constructs bookmark collections of reasonable size for a group's common needs which are represented as topics. The key points are identification of the topics out of individual users' interests, and during the construction procedure taking users' personal view of the relations between bookmarks into consideration. This is dependent on the specific user model and group model of applications. For instance, in LiveMarks, users express their interests in the form of "groups-tasks-queries-bookmarks". In the collaborative applications like BSCW[2], users are able to formulate queries, review results and organize the information into folder-based shared workspace. Thus, the representation of users' interests varies with different user interface. From the retrieval point of view, however, they can be generalized as presented in the model. The denotation can be lists of descriptors. In such lists each descriptor carries a weight indicating its relevance with the particular interest.

A topic is then defined as the shared interest within a group or across groups. It can be identified by clustering user interests or text mining approach. Topics are dynamically created and updated according to the changes in users' interests which are also captured dynamically, e.g. based on a periodical schedule or on users' demand. Both user interests and topics are made uniform by structured descriptor lists or thesauri, or using an ontology. For each topic, a user group is defined associating users to topics so as to form a basis for retrieval. It could also be used to recommend users with similar interest by calculating similarity of user interests, or to support collection-mediated collaboration.

3.4 Bookmark collection and repository

For each topic, relevant web pages are collected to form a topic-specific collection. An agent-based crawler generates a set of queries out of the topic's description and initiates gathering process. It wraps meta search like search engines to acquire preliminary materials. From a recommender service, it explores socially constructed knowledge to obtain highly-rated URLs. Then it performs link following to fetch the original web pages. These Web pages are filtered by a wide range of criteria, e.g. spamming, semantic redundancy, hostname alias, syntax error, , robot exclusion protocol, etc. Additional information concerning a Web page such as hyperlinks, meta content and lastModifiedDate, etc., are extracted as its properties to enrich its description based on the statistics introduced in [1].

Each fetched Web page is parsed into a uniformed text vector of <descriptor, weight> and internally represented as a bookmark with properties like title, excerpt, lastModified, etc. The weight of a descriptor is derived based on the relative term occurrence frequency. A bookmark is then created for each Web page. The definition of bookmark is intended to be compatible with standards of web content such as MCF and RDF [5]. It is classified regarding existing topics or can be used to generate new topics. In this way each topic gradually accumulates its own collection of bookmarks with varying relevance in content. The frequently appeared descriptors within the bookmarks are extracted and put into the topic's descriptor lists to extend the topic description.

All the collected bookmarks are stored within a single repository. The repository deploys an efficient storage scheme and develops maintenance services to guarantee the validity of the bookmarks. A bookmark storage and access architecture is designed to support database-transparent access of the bookmarks collections in an object-oriented way, and to enable interoperation between repositories, e.g. exchange of bookmarks.

3.5 Indexing, classification and retrieval

To normalize the vocabulary used to express users' interest, topics, and bookmarks, published glossaries, e.g. ACM Computing Reviews Dictionary, are taken as standard termlist. A larger scope of vocabulary for queries and the indexed bookmarks is supported by extracting the words from the collection itself to construct topic-specific thesaurus. Collected bookmarks are stemmed and indexed by the same set of descriptors. A varied inverted file organization is constructed for the descriptors, i.e. each descriptor in the dictionary is related to a group of relevant topics each of which is associated to a group of bookmarks. In this way those bookmarks that don't contain the descriptor still can be recommended as relevant documents for a particular user's queries, whereby user s are able to get good results even for poorly-formulated queries. Thus a better recall/precision value is achieved. Similarly hyperlink indexes are also generated to support full-fledged retrieval.

To provide better performance, a topic can be splited and the corresponding bookmark collection can be reclustered. The classification scheme is based on statistical term association [16]. All the descriptors are compared and high frequency keywords are extracted either to extend the topic or to identify a new topic. Each bookmark is then analyzed to determine which topic it belongs to. The classification can be overlapped, i.e. one bookmark may belong to multiple collections. On the other hand, bookmarks can be classified based on the form rather than the content, e.g. entry page, bookmark list page in order to provide user with an option of intended Web page types.

When a user issues a new query, it is parsed into vector model-based representation and matched with collected bookmarks according to a two-step retrieval scheme. It firstly matches with relevant topics then the bookmarks within the collection designated by the topic. The contextual information of a query is also delivered to match more topics for which the user hasn't register yet. The matching algorithms use the correlation coefficient, e.g. cosine function or distance measure [16].

3.6 Coordinated agents for TopicMark

Since social agent technology[17] is very fit for the autonomous work needed by the TopicMark system, we decide to select it as the base technology for implementation.

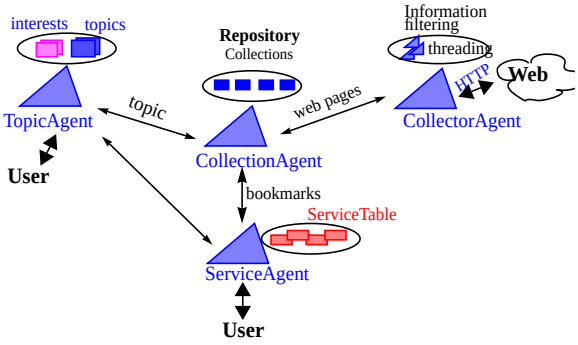

Figure 4. Agents' interaction in TopicMark.

As figure 4. shows, the agents in TopicMark are specialized with regard to behavior and function. Each type of agent uses the knowledge of its environment and has specialized task to accomplish. They plays different roles in the overall system. For instance, TopicAgent is designed to formulate and manage topics and it cooperates with CollectionAgent in order to build collections for particular topics. They interact with each other by exchanging messages of certain types ("performative"). The interaction- synchronous or asynchronous--can be conceived of as a conversation. Such conversation patterns may be formalized as finite-state-machines or in a distributed environment as high-level Petri nets [19]. The FSM specification of each conversation can be described by conversation tables which specify the state transitions as well as the message transferred in particular states. They may be used to formally verify that the conversation is free of deadlock even in the presence of message delay and mixed initiative of the conversation partner (not a simple turn-taking protocol). In addition their activities are monitored by a coordinator agent (not shown in the figure). It synchronizes the state of TopicMark agents so as to ensures the reliable and predictable behavior of the whole system.

4. TopicMark implementation

As a service allowing groups to accumulate topics and bookmark collections, TopicMark is implemented on top of an agent open infrastructure [18] as an application service. The infrastructure is a distributed platform for interacting social agents providing a runtime environment with system functionality such as an agent naming/addressing schema, an agent messaging mechanism, a directory service, etc. This platform consists of the kernel modules–an agent engine and system communication services. Both TopicMark and the platform are implemented in Java.

TopicMark is implemented as a domain-independent service. Various collaboration system can integrate it by following the protocols specifying the user interest and retrieval interface. It has basic architectural support such as request scheduling, transaction process and persistence support. By distributing functionality over network, the scalability of the service is expected to be guaranteed. In addition the information exchange protocol between individual TopicMark system and other information service such as a recommender service like LiveMarks, is specified to support the system's interoperability.

The first prototype is developed and experimented as a service in LiveMarks within our research group at GMD. It includes agents which manage topic formulation, build bookmark collections, and mediate service requests. This prototype is intended for demonstration, exploratory use, and evaluation in cooperation with an industrial partner from the oil business. Prospective users are members of project teams operating in oil field development. Team members with different background respectively belong to special professional groups. They are usually spread around the world, and may belong to several teams at the same time. Information retrieval and exchange is central to their work. Construction of indexes of the Web resources relevant to their interest and preference is essential to meet their needs and facilitate the information gathering process.

5. Conclusion

This paper presents a user-centric approach for collaborative information gathering. An interest-topic model is described briefly. It focuses on exploiting the shared aspects of users' work. Base on the model, a TopicMark system is developed to support collaboration within/across groups through information sharing. The presented approach represents a novel way for resource discovery on the Web in order to meet people's information needs within various collaborative environment.

Although TopicMark represents a possibly promising solution for collaborative information gathering, there are some important issues to investigate in order to build a useful information service. First of all, the personal aspects of users' interest might be lost in the topic-construction process and make it difficult to support users' information gathering as individuals. Building personalized topic structure with regard to each user's preference is a possible solution for this problem. Secondly, since users may use a different vocabulary to express their interests, it is critical to integrate these vocabularies in order to ensure topics to have an unambiguous meaning.

The distributed bookmark services presented in this paper form a network of topic-focused bookmark collections and emerge as a valuable information resource for users to access through popular Web browsers. In the future, we propose to integrate TopicMark service with BSCW shared workspace

system [2] to achieve more understanding within extensive collaborative contexts. Some important issues to be investigated include user feedback, topic-reformulation, media migration and so on.

6. References

[1]Allison Woordruf, et al. "An Investigation of Documents from the World Wide Web," in Proc. 5th Int. WWW Conf. May 6-10, 1996, Paris, France.

[2] Bentley, R., Appelt. W., Busbach, U., Hinrichs, E., Kerr, D., Sikkel, K., Trevor, J., Woetzel, G. "Basic support for cooperative work on the World Wide Web," Int. J. Human Computer Studies 46 (1997), 827-846.

[3]C. Mic Bowman, Peter B. Danzig, Darren R. Hardy, Udi Manber and Michael F. Schwartz "The Harvest Information Discovery and Access System," Computer Networks and ISDN Systems 28 (1995), 119-125.

[4]Daniel Tkach, "Text Mining - Turning Information Into Knowledge, A White Paper from IBM". http:// www.ibm.com. Sept, 1998.

[5]David Singler, et.al. "Resource Description Framework (RDF), "WD-RDF-Schema-1998011, W3C 1998.

[6]Eui-Hong(Sam) Han, Daniel Boey, Maria Gini, Robert Gross, Kyle Hastings, George Karypis, Vipin Kumar, Bamshad Mobasher, and Jerome Moore, "WebACE: A Web Agent for Document Categorization and Exploration," Proceeding of Agents'98. 1998.

[7]FOLDOC-Free On-Line Dictionary of Computing, http:/ /www.easynet.de/resources/foldoc/index.html

[8]Guo, H., Kreifelts, Th., Voss. A. "SOAP: Social Filtering through Social Agents," ERCIM Workshop report No.98/ W001. 5th Int. DELOS Workshop: on Filtering and Collaborative Filtering Budapest, Hungary, 10-12 November, 1997.

[9]J.O. Pedersen, C. Silverstein, C.C. Vogt (Verity, Inc.) "Verity at TREC-6: Out-of-the-Box and Beyond,", Overview of the Sixth Text REtrieval Conference (TREC-6). 1997.

[10]Julian A. Yochum, "Research in Automatic Profile Generation and Passage-Level Routing with LMDS", In D. Harman, eidtor, *Overview of the Third Text REtrieval Conference (TREC-3)*, pp-289-297.1995

[11]Liren Chen, Katia Sycara, "WebMate: A Peronal Agent for Browsing and Searching,", Proceeding of Agents'98. 1998.

[12]Peter Pirolli, Patricia Schank, Marti Hearst, Christine Diehl, "Scatter/Gather Browsing Communicates the Topic Structure of a Very Large Text Collection," Proceeding of CHI'96. 1996.

[13]Resnick, P., Varian, H.R. "Recommender Systems, " Comm. ACM 40, 3(1997), 56-58.

[14]Richard M. Keller, et.al, "A Bookmarking Service for Organizing and Sharing URLs,", Proceeding of Sixth International World Wide Web Conference. 1997.

[15]Robert B. Allen, "User models: theory, method, and practice", Int. J. Man-Machine Studies 32 (1990), 511-543.

[16]Salton, G. "Automatic Text Processing", Addison-Wesley, Reading, MA, 1989.

[17]Voss, A., Kreifelts, Th. "SOAP: Social Agents Providing People with Useful Information," in Proc.Int. ACM SIGGROUP Conf. on supporting group work, ACM, New York NY, 1997, pp-291-298.

[18]Voss, A, Guo, H, Hausen, H.-L., Juhnke, M, Nakata. Kreifelts, Th., and Paulsen, V.(1998) "Agents for Collaborative InformationExploration", in Proceeding of the Third International Workshop on CSCW in Design (CSCWID'98), Tokyo (to appear).

[19]Woetzel, G., Kreifelts, TH. "Deadlock freeness and consistency in a conversation system," in B.Pernici, A. A. Verrijn-Stuart (eds) Office Information Information Systems: The Design Process, Proc. IFIP WG 8.4 Work. on Office Information Systems: The Design Process, North-Holland, Amsterdam, 1989, pp.239-253.