Supporting Federated Information Sharing Communities

Bicheng Liu
Smart Web Technologies Centre
The Robert Gordon University
Aberdeen, United Kingdom
b.liu@rgu.ac.uk

David J. Harper
Smart Web Technologies Centre
The Robert Gordon University
Aberdeen, United Kingdom
d.harper@rgu.ac.uk

Stuart Watt
School of Computing
The Robert Gordon University
Aberdeen, United Kingdom
s.n.k.watt@rgu.ac.uk

ABSTRACT

In this paper we describe the concept of Federated Information Sharing Communities (FISC), and associated architecture, which provide a way for organisations, distributed workgroups and individuals to build up a federated community based on their common interests over the World Wide Web. To support communities, we develop capabilities that go beyond the generic retrieval of documents to include the ability to retrieve people, their interests and inter-relationships. We focus on providing *social awareness* "in the large" to help users understand the members within a community and the relationships between them. Within the FISC framework, we provide *viewpoint retrieval* to enable a user to construct visual contextual views of the community from the perspective of any community member. To evaluate these ideas we develop test beds to compare individual component technologies such as user and group profile construction and similarity matching, and we develop prototypes to explore the broader architecture and usage issues.

Categories and Subject Descriptors: H.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.3 [Information Storage and Retrieval]: Online Information Services

General Terms: Design, Experimentation

Keywords: Information sharing community, social awareness, viewpoint retrieval, rational link

1. MOTIVATION

Increasingly, the World Wide Web is being viewed as a means of creating web communities rather than simply as a means of publishing and delivering documents and services. We characterise current approaches to building web communities as communication-centred, document-centred or community-centred. Communication-centred approaches focus on communication such as chat, presence, etc [2]. Content-centred approaches define communities as collections of documents sharing a common topic interest [1]. In the community-centred approaches, the focus is on building an understanding of community relationships through social awareness [3], and this is the main thrust of our work.

We have developed the concept of FISC, and associated architecture, which we believe provides a natural and effective way for organisations, distributed workgroups and individuals to build up federated communities based on their common interests. Prominence is given to overviews of the community that show important relationships between community members, based on *rational links* and presented through viewpoints structured according to the community structures. Within this framework, we provide viewpoint retrieval that enables a user to view the community from the perspective of any other part of the community. Communication facilities also provide personalized notification of changes to the community, and communication between members. FISC leverages organisational and social structures with document content to provide a community-centred information sharing and communication environment.

The scenario in Figure 1 brings out some of the more detailed features of a FISC-based community from the user's point of view.

> Jamie is a new research student, she is a member of information retrieval (IR) group, and is working on "user interfaces and environments for collaborative information seeking". As the first step of her study, Jamie wants to get some relevant information about the topic and understand the structure of the IR community.
>
> Jamie logs into the "IR Europe" FISC community. She issues query "user interfaces and environments for collaborative information seeking" and chooses to use the community-viewpoint for a contextual overview of the topic. Rather than returning a long list of relevant documents, FISC identifies for her:
>
> - Main IR groups working in this area - IR/A and IR/B
> - A very relevant project plus associated individuals - P/C
> - An individual working in human computer interaction (HCI) - I/D
>
> Jamie asks the system to track these for new publications, new projects, and new personnel, possibly filtered according to the query, and alert her to changes.
>
> Jamie then views her group, and explores other groups, projects etc. through rational links. In this way she understands her group in relation to the rest of the IR community in Europe. This view can, if require, be refined by a query.
>
> Jamie decides to view this world through some other viewpoints, such as those of groups IR/A and IR/B, project P/C and individual I/D. She discovers new relevant projects etc in this way. Importantly, the I/D viewpoint leads Jamie into another related FISC community on 'HCI in Europe'.

Figure 1: Scenario of a FISC-based community

In the rest of this paper, we outline the design of the FISC architecture, and introduce the evaluation strategies including experiments based on the "Reuters Collection", and the Web Network and "CiteSeer Community" prototypes.

SIGIR'04, July 25–29, 2004, Sheffield, South Yorkshire, UK.
ACM 1-58113-881-4/04/0007.

2. DESIGN OF FEDERATED INFORMATION SHARING COMMUNITIES

The foundation of FISC is to help people find out and understand the state, activities and relationships of community members, i.e. support discovering of awareness within a community. These requirements are reflected in the three-layer FISC architecture, as shown in Figure 2.

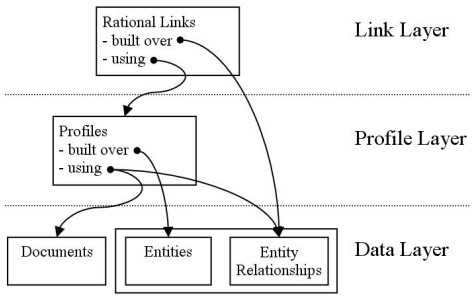

Figure 2: The FISC outline three-layer architecture

The information layer at very bottom of the FISC architecture contains the basic elements of a community - documents, entities (e.g. individual, group, project) and entity relationships (e.g. containment relationship between individual and group). Generally speaking, each community entity will have an associated set of documents. FISC represents each entity by one (or more) profiles based on the content of the entity's associated set of documents. Using these profiles, we create links between entities as representation of their relationships in the link layer. Potentially, there are many different types of links (relationships) within a community, including content-based links, organisational links and social links. Our work focuses on content-based links in order to provide a content-based relationship structure within the community, and use such information to deliver/derive the awareness of "who is working on what" and "who is working with whom".

Given a specific user's awareness request, it is important to identify which are the most valuable links - we define these as rational links, and present this information through viewpoint retrieval. A viewpoint is a set of rational links from the perspective of a given entity, which are presented in a view structured according to the specific request (bearing in mind that viewpoints can be further refined through queries and filters). For example, we could provide a viewpoint for an individual, showing rational links to groups, or individuals from different groups. Moreover, viewpoint retrieval allows users to construct a visual contextual overview of the whole community from the community viewpoint.

3. EXPERIMENTATION & PROTOTYPING

In order to implement a FISC-based system to test and validate the concept, we need to develop/devise appropriate techniques underpinning the architecture sub-components. A range of techniques has been selected through literature survey, which include language modeling for profile generation, document clustering and similarity measurement for autonomous content-based linking.

We propose using the "Reuters Text Categorization Test Collection" to explore the effectiveness of the above-mentioned IR techniques. We will construct pseudo-groups by randomly partitioning each category and create links between the groups. All pseudo-groups being derived from the same category are assumed to be connected by rational links. For example in Figure 3, rational links are between groups G1, G2 and G3 generated from category C1, and between groups G4 and G5 from category C2. Given this setting, we will experiment different techniques for identifying these rational links. More complex experiments are envisaged to explore linking of multi-category groups, and to investigate the robustness of linking when "noise" documents are added to groups.

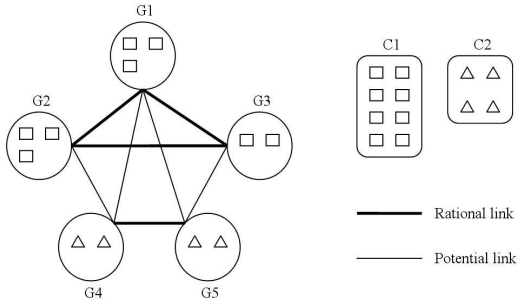

Figure 3: Experimentation using Reuters Collection

We have designed the Web Network and "CiteSeer Community" prototypes to test and evaluate the FISC architecture. The Web Network prototype generalises the concept of Web Ring, and is an instantiation of the FISC architecture. Through building this prototype, we will develop the underlying FISC software infrastructure (e.g. application interface). The summative evaluation of FISC will be based on the "CiteSeer Community" prototype. This prototype addresses the weakness of the scientific literature digital library CiteSeer, in supporting users in finding out activities and relationships of community members. It represents CiteSeer as a FISC community with a sensible structure, and provides an ideal platform for qualitative user experiments.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have described the concept of FISC and outlined its architecture for supporting information sharing communities over the World Wide Web. Notions such as profile generation, rational link and viewpoint retrieval have been identified as new challenges in our research.

In the next step of work, we will identify the appropriate algorithms and techniques for implementing the Web Network and "CiteSeer Community". The design of FISC architecture will be further investigated and revised through experiments. Future work will include research on core retrieval representations and algorithms, social network analysis, user interface design and field evaluation.

5. REFERENCES

[1] G. W. Flake, S. Lawrence, C. L. Giles, and F. Coetzee. Self-organization of the web and identification of communities. *IEEE Computer*, 35(3):66–71, 2002.

[2] S. U. Kelly, C. Sung, and S. Farnham. Designing for improved social responsibility, user participation and content in on-line communities. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 391–398. ACM Press, 2002.

[3] O. Liechti. Awareness and the www: an overview. *ACM SIGGROUP Bulletin*, 21(3):3–12, 2000.