Collaborative Ad-hoc Information Sharing in Cross-Media Information Environments

Beat Signer[1], Alexandre de Spindler[2], and Moira C. Norrie[2]

[1] Vrije Universiteit Brussel
Pleinlaan 2
1050 Brussels, Belgium
bsigner@vub.ac.be

[2] Institute for Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
{despindler,norrie}@inf.ethz.ch

Abstract. Due to the ever increasing number of different digital media types that we use in our daily work, it is no longer sufficient to manage them in an isolated way but desirable to define associations across the media boundaries. While cross-media information systems allow different forms of media objects to be linked, there is often a lack of support for the flexible authoring and sharing of these links. We present a solution for exchanging cross-media link information in an ad-hoc way among communities of users based on a notion of a cooperative cross-media link server. We describe how the system has been implemented based on an existing cross-media link server (iServer) and peer-to-peer technologies.

1 Introduction

It is common nowadays for user communities to want to create, access and share heterogeneous collections of associated multimedia objects. While hypermedia systems, and specifically cross-media link servers, allow different types of media objects to be linked together in flexible ways, often there is a lack of support for flexible means of authoring and sharing these links.

Our interest was to support dynamic and temporary forms of user communities where users come together in physical or virtual space to support each other in a specific task. For example, a group of designers attending a meeting and a group of students attending a course can both be thought of as dynamic and temporary user communities. Typically, these user communities will want to create ad-hoc and transient forms of collaborative information spaces that support a particular activity such as planning a project or learning for an exam. It should therefore be possible to share link data in an ad-hoc way to support the community-based authoring of links between resources that define such transient collaborative information spaces.

We note that allowing users to exchange information in a peer-to-peer (P2P) manner rather than via a central server offers the potential for more dynamic forms of information sharing. For example, in the case of a design meeting, information could be exchanged directly between personal devices such as laptops and also between these and fixed devices such as an interactive table.

Our solution was to develop a cooperative cross-media link server for exchanging link information in an ad-hoc way among communities of users. It is important to note that, in our solution, link services are discovered in a completely decentralised manner. Users can dynamically join the community and start annotating or adding links to arbitrary third-party resources as well as access link metadata from other users. To avoid issues of information overload and link fraud, we introduce a collaborative filtering mechanism based on a combined ranking of users and link resources.

In this paper, we describe how the system was implemented based on an existing cross-media link server, called iServer [1], and P2P technologies. We start in Sect. 2 with a more detailed look at the motivation for collaborative cross-media information spaces and a discussion of related work. In Sect. 3, we briefly introduce the existing iServer platform and its underlying cross-media link model before going on to describe the functionality and operation of the cooperative iServer version. Different approaches for user and link rating are discussed in Sect. 4. Concluding remarks are given in Sect. 5.

2 Motivation

The original hypertext and hypermedia information models are based on the concept of connected document spaces in which additional meaning is usually associated with the links between documents. Early models have been extended over the years, both to broaden the scope of the model and to improve the functionality and maintenance of systems through more powerful and flexible link management. With the emergence of ubiquitous and pervasive computing, physical hypermedia systems have been proposed that enable real-world objects to be linked to digital media, and vice versa, by allowing physical resources to also be included as nodes in the connected information space [2].

Underlying all of these developments, the basic information model remains the same and an information space is defined by a connected graph where the nodes are resources and the links are represented by edges. The anchor and target of a link can either be an entire resource or an element within a resource. We have realised a general framework to support a range of hypermedia tools and services by adopting a database approach to the problem that first involved developing a generic link metamodel and then implementing it using an object database management framework.

A benefit of managing links separately from resources is that new links can easily be created by arbitrary users. In these *open link authoring* systems the information space evolves over time based on the users' current interest. For example, in a teaching and learning environment, students can not only consume the material prepared and published by a teacher, but also add their own links between different resources (e.g. slide handouts, web pages and PDF documents).

Links can not only be created between existing resources but also associations to content that has been created during a lecture can be defined. Examples of solutions for capturing notes during a lecture or meeting and sharing them online include the Pulse Smartpen application from Livescribe[3] or the PaperPoint [3] interactive paper presentation manager.

When it comes to link sharing, students will no doubt find that certain students provide more useful links than others. A course will evolve over time and information of fellow students might be more reliable than information added by former students. On the other hand, information of previous students who obtained excellent grades might also be considered as especially valuable. The introduction of a notion of link quality based on user trust helps to deal with these issues. Our open link authoring approach combines the traditional publisher and consumer model with a democratic authoring process based on an open link platform that supports existing and emerging types of resources

The teaching environment is only one application scenario where people want to organise and share information with other local or remote users. Our iServer solution is not limited to a specific application, but rather provides a general platform for cross-media information management that can be applied to a large variety of application domains. The community-based link sharing, as provided by the presented P2P iServer framework, is related to proposals in distributed hypermedia [4]. A benefit of our approach lies mainly in the means to realise the goal. By building on advanced database technologies and exploiting practices of metamodel-driven system engineering to the full, we have succeeded in designing a flexible link server platform that supports digital and physical resources and is open to new communication paradigms including P2P technology.

3 Cooperative Cross-Media Link Server

iServer is a cross-media link server capable of supporting an extensible set of digital and physical media. The framework is based on the resource-selector-link (RSL) metamodel and supports the integration of new media types based on a resource plug-in mechanism. It is beyond the scope of this paper to describe all details of the RSL model but the complete model can be found in [5].

The RSL model consists of three core *entity* types represented by *resources*, *selectors* and *links* as highlighted in Fig. 1. A link is associated with one or more source and target entities where each of these entities might be either a resource, a selector or another link. While a resource represents an entire information unit or service (e.g. a web page or a PDF document), a selector allows us to address parts of a resource. Note that a selector always has to be associated with a single resource but a resource may have multiple selectors. By modelling links as an entity subtype, we gain the flexibility to define links with other links as source or target entities.

The user model defines the access rights associated with an entity and is essential for information sharing issues discussed later in this paper. Users may

[3] http://www.livescribe.com

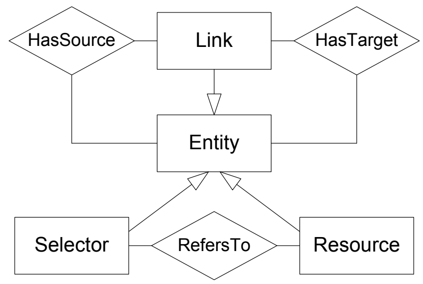

Fig. 1. Resource, selector and link

be either *individuals* or *groups* and each entity has exactly one individual creator who defines the user access rights. When it comes to sharing links and resources in the P2P version, it is up to the creator of an entity to specify whether the information is private or should be publicly available.

While the RSL model defines the abstract concepts of resources and selectors, the system can be extended to support new media types by providing concrete plug-in implementations for the resource and selector types. In the past, various iServer plug-ins have been developed to support different media types. This includes the iWeb plug-in for HTML resources as well as resource plug-ins for images, sound and movies or an RFID plug-in for tagging physical objects. Furthermore, a plug-in for linking paper documents has been realised based on interactive paper (iPaper) technologies [1]. A number of applications have been implemented based on the iServer framework, including generic cross-media browsers, a paper-based mobile tourist information system [6] as well as educational applications and information services. Based on our experience, the authoring of links in such highly-connected cross-media information spaces can be time-consuming and tedious. By making the cross-media authoring tools available to the user, we move from the publisher and consumer model to an open authoring system where every user becomes a potential publisher and may share their link information with other users.

In our distributed iServer solution, a client always accesses a resource from its original location and, only in a second step, is additional link metadata acquired over the dynamic P2P iServer network as illustrated in Fig. 2. The separation of content and metadata implies that a resource should always be available, provided that the server on which the resource is hosted is up and running, whereas any additional link information may change dynamically over time based on the set of iServer peers currently available in the network.

We have to distinguish between *persistent link metadata* in the form of link information that is stored in a personal iServer instance and the *transient link metadata* received from the set of remote iServer peers. While the quality of the persistent link metadata can be ensured by controlling the users who have access to a personal iServer instance, we do not have any direct control mechanism to guarantee the integrity of information provided by remote peers. Therefore, an

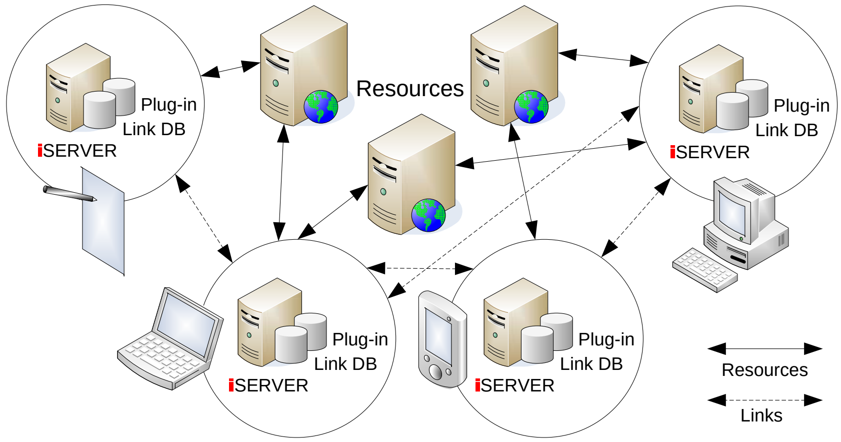

Fig. 2. Cooperative iServer architecture

essential part of the cooperative iServer framework is a collaborative filtering mechanism based on the rating of remote users and links.

Of course, we are not alone in this view since distributed link services, where link information is stored separately from the resources, have investigated associative linking of resources. More recently, P2P technologies have been considered for building new forms of open hypermedia systems where central link databases are no longer required [7]. Our approach differs in several aspects from other P2P-based open hypermedia systems. First, we do not use the P2P network functionality to replicate any hypermedia documents and perform searches in the distributed network of peers as proposed by Larsen and Bouvin [8]. In our solution, the original hypermedia documents are clearly separated from any annotations or external links and it is only the link or annotation metadata that is shared in an ad-hoc manner. In contrast to other systems, our approach is not based on finding resources which are similar to the resource at hand, but rather on associations between resources that users themselves have found to be meaningful which is closely related to the idea of associative linking originally proposed by Bush [9].

In designing a distributed architecture for iServer, it was one of our goals to ensure that the framework kept to its principle of being as general as possible, with a clean separation of concerns. Even if iServer is primarily an extension of an object database system, it provides well-defined Java, XML and Web Service APIs for accessing and updating information. The general interaction between peers consists of sending single API calls from one peer to another, the execution of that request on the remote site and the transfer of the result back to the initiating peer. The information returned by the remote peers has to be combined and integrated with information that is available from the local iServer instance. The functionality of a remote iServer system is offered to a local iServer instance via a peer service. More details about the P2P-based distribution architecture can be found in [10].

4 User and Link Rating

In a cooperative community of publishers and consumers who are equally responsible for the available content, a variety of issues must be addressed. Communities where participants consume without contributing produce little content and result in a heavy load on a few nodes. A community may also be a target for the distribution of unsolicited or malicious content. Some systems try to avoid the lack of contribution by keeping designated publishers to ensure the information supply. The content in our information spaces consists of resources and links and we primarily envisage shifting the authoring of links between resources to the user community. The effort required from the user should be minimal to encourage them to contribute.

The rapid and considerable emergence of technologies enabling democratic publishing has led to the development of rating and filtering systems addressing the issues of unsolicited or malicious content. Existing filtering techniques such as content-based or collaborative filtering have in common the fact that the rating inference is based on some notion of similarity [11, 12]. However, the extraction of relevant features used for comparing resources, the computation of similarities and the cold start problem still form a bottleneck in current systems.

In real world situations, the filtering of resources is often based on the trustworthiness of the provider in addition to the relevance and quality of the resource. Electronic communities such as eBay[4] have successfully implemented trust and reputation techniques supporting users in taking decisions. Research has given rise to a variety of techniques for the rating of users and the interpretation of rating values [13, 14]. In all these cases, a set of user ratings can be represented by a *rating graph* where a directed weighted edge connects a rating user with a rated user. If two users are not connected by an edge, there possibly exists one or multiple paths connecting them transitively. Therefore, a user who knows a relatively small number of other users and is able to rate them can infer ratings from a vast number of other users by a transitive propagation of ratings.

Our framework supports individual users in deciding on their own levels of trust in other users. We provide a rating implementation where a rating value is interpreted as the amount of trust flowing from a rating user to a rated user. The transitive propagation of trust consists of searching the minimum trust along a path connecting the two users. If multiple such paths exist, all minima are summed up. This can be achieved by running a maximum flow algorithm on the rating graph with the rating user as a source and the rated user as a sink.

The main interaction of iServer peers consists of exchanging link metadata. A common request is a query for in- and outgoing links of a particular resource. As a result, the user receives a set of links from multiple users. Filtering can

[4] http://www.ebay.com

be achieved by deciding, for each item received from a particular user, whether to accept or reject it. The information available for this decision consists of the currently received resource item, the sending user and the set of items previously received from other users. The currently received item can be rated in the context of the previously received ones, for example by computing its frequency. Therefore, we combine user ratings with response ratings to filter responses returned from remote peers. This may, not only help to improve the quality of information presented to the user, but also reduce the quantity, thereby preventing the information overload that could result from a large, highly-connected information space.

We encode user ratings as tuples containing the rating user, the rated user and the rating value. Such a tuple is created by the rating user and propagated to all other members of the peer group. A user rating manager ensures that all peers have the same set of tuples stored locally. The tuple synchronisation is achieved as follows:

1. On startup, the peer reads a file containing tuples stored in previous sessions. If this file does not exist, a new empty one is created.
2. The peer creates a tuple set S_{local} containing all tuples in the file. Whenever S_{local} changes, the file is updated.
3. When a peer joins the group, it requests the tuple set S^i_{remote} from all other members i of the group.
4. Every incoming tuple set S^i_{remote} is compared with the local set S_{local} as follows:
   – If S_{local} does not contain all tuples in S^i_{remote} then the local set is updated.
   – If S^i_{remote} does not contain all tuples in S_{local} then the local set is broadcast to all other members.
5. When no more sets are broadcast, all tuple sets contain the same tuples.
6. Whenever a new rating is set locally, the local set of tuples is broadcast to all other members of the group.

Every iServer P2P API request is broadcast to all members of the groups. As a result, a requesting peer possibly receives multiple responses. A response rating manager filters incoming responses before they are made accessible to the user. The selection is based on rating values that are computed by response raters for every response. The rating values of multiple response raters are combined by an aggregator function. We implemented two response raters, one returning the rating of the responding user as described above and the other returning the frequency of the response within the collection of previous responses. Further details about the link rating and the visualisation of shared links as well as applications that are based on the presented framework can be found in [10].

5 Conclusion

We have presented a notion of collaborative cross-media information environments based on community-based link authoring between arbitrary resources and elements within these resources. Cooperation is supported on two levels—the user management integrated into the core RSL model and a P2P architecture for distributed information sharing. In order to avoid information overload and ensure link quality, we introduced a collaborative filtering mechanism that is based on a graph model. A rater has been presented that implements filtering using a maximum flow algorithm. We are currently evaluating the similarity measure in the filtering process described in this paper which involves the implementation of a third response rater and its deployment within our iServer P2P framework.

References

1. Signer, B.: Fundamental Concepts for Interactive Paper and Cross-Media Information Spaces. PhD thesis, ETH Zurich (May 2006) Dissertation ETH No. 16218.
2. Hansen, F.A., Bouvin, N.O., Christensen, B.G., Grønbæk, K., Pedersen, T.B., Gagach, J.: Integrating the Web and the World: Contextual Trails on the Move. In: Proc. of Hypertext 2004, Santa Cruz, USA (August 2004)
3. Signer, B., Norrie, M.C.: PaperPoint: A Paper-Based Presentation and Interactive Paper Prototyping Tool. In: Proc. of TEI 2007, Baton Rouge, USA (February 2007)
4. DeRoure, D., Carr, L., Hall, W., Hill, G.: A Distributed Hypermedia Link Service. In: Proc. of SDNE '96, Washington, USA (1996)
5. Signer, B., Norrie, M.C.: As We May Link: A General Metamodel for Hypermedia Systems. In: Proc. of ER 2007, Auckland, New Zealand (November 2007)
6. Belotti, R., Decurtins, C., Norrie, M.C., Signer, B., Vukelja, L.: Experimental Platform for Mobile Information Systems. In: Proc. of MobiCom 2005, Cologne, Germany (August 2005)
7. Bouvin, N.O.: Open Hypermedia in a Peer-to-Peer Context. In: Proc. of Hypertext 2002, College Park, USA (June 2002)
8. Larsen, R.D., Bouvin, N.O.: HyperPeer: Searching for Resemblance in a P2P Network. In: Proc. of Hypertext 2004, Santa Cruz, USA (August 2004)
9. Bush, V.: As We May Think. Atlantic Monthly **176**(1) (July 1945)
10. Signer, B., de Spindler, A., Norrie, M.C.: A Peer-to-Peer-based Distributed Link Service Architecture. Technical Report TR636, ETH Zurich (August 2009)
11. Claypool, M., Gokhale, A., Miranda, T., Mumikov, P., Netes, D., Sartin, M.: Combining Content-Based and Collaborative Filters in an Online Newspaper. In: Proc. of the ACM SIGIR Workshop on Recommender Systems, Berkeley, USA (August 1999)
12. Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J.T.: Item-based Collaborative Filtering Recommendation Algorithms. In: Proc. of WWW10, Hong Kong, Hong Kong (May 2001)
13. Jøsang, A., Ismail, R., Boyd, C.: A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems (2005)
14. Gambetta, D.: Can We Trust Trust? In: Trust: Making and Breaking Cooperative Relations. Department of Sociology, University of Oxford (2000)