Recommendation in a Collaborative Digital Library Environment

Norbert Fuhr*, Norbert Gövert, Claus-Peter Klas
University of Dortmund, Germany

Federated digital libraries aim at covering most of a user's information needs. The DAFFODIL system[1] developed by our group is an agent-based architecture for accessing a relevant set of digital libraries in the area of computer science. DAFFODIL integrates searching and browsing, supports high-level search activities and allows for horizontal and vertical integration of services.

In order to support end users effectively, a digital library system should be integrated in the end user's work as much as possible. For this reason, we have started work for combining DAFFODIL with a groupware service that allows for storing digital library objects (DLOs), for annotating them and for sharing them with other users. Since we target at a large number of users, we want to include a recommendation service that suggests digital library objects to users or groups and that also supports the formation of user groups with similar interests.

In the following, we first briefly describe the collaborative digital library environment and then discuss recommendation issues.

DAFFODIL supports a number of DLOs such as bibliographic records as well as the fulltext of documents, references journals, conferences, authors and research groups. Each of these objects is assigned a unique identifier within DAFFODIL, which is achieved by means of a central object repository that stores descriptions of all objects currently referenced by any user of the system.

The groupware service allows a user to store DLOs in a hierarchy of folders. Users may form groups, either as closed project teams or open communities and share folders within these groups. (Communities and their folders are visible for users outside of this group, whereas project teams are not). Objects stored in the groupware service may be annotated by one or more users by attaching a discussion thread to them.

In this collaborative setting, the task of the recommendation service is to suggest DLOs to users or groups and to support the formation of new groups. Instead of relying on users' rating (empirical evidence from similar applications suggests that only little rating activity should be expected), the major basis for recommendation will be the organization of DLOs in folders. The general approach for recommendation is the comparison between users and/or groups based on their stored sets of DLOs. If there is sufficient overlap between the sets, the remaining objects can be recommended to the other users/groups not referencing these objects yet. In contrast to the standard recommendation procedure, however, there are two major differences:

1. Whereas standard applications consider one type of objects only, the data set forming the basis of the recommendation procedure consists of a variety of object

*fuhr@cs.uni-dortmund.de

[1] `http://www.daffodil.de`

types.

2. There are different kinds of relationships between DLOs, which should be taken into account when comparing users' profiles.
3. Usually, the data belonging to a single user forms a single set. In our setting, users organize their data in different folders, which are representing different contexts.

For the first problem, we envisage the assignment of different weighting factors for different object types (e.g. if two users reference the same article, this may not be as significant as referring to the same author). Vice versa, for the suggestion of new objects to a user, there may be different threshold values depending on the type of the object.

When comparing sets of DLOs, different kinds of similarity or relationships between objects should be considered:

- Similarity based on content: For documents, we usually have some metadata characterizing their content; for the other types of objects, however, the construction of appropriate content descriptions may be computationally expensive (e.g. by aggregating the descriptions of all documents published by an author or in a journal).
- Structural relationships between different object types: Here we mainly have author and aggregation relationships (e.g. document – journal, document – conference, author – research group). As a straightforward approach for considering these relationships in the recommendation procedure, related objects could be given a lower weight than identical objects.
- Relationships between objects of the same type. Most of these relationships are inferred from the author and aggregation relationships (e.g. co-authorship, co-occurrence of documents in a journal issue). In contrast, citation links are given explicitly. Again, different weighting factors could be used for considering these types of relationship.

The folder structure should be considered by the recommendation process, in order to distinguish between the different interests of a single user. Thus, instead of taking into account all DLOs referenced by a user, only objects within the same folder should be regarded for recommendation. However, since users may follow different strategies in organizing their folders, the size of a folder may vary considerably, thus making comparisons rather difficult. As a simple strategy, we plan to fix a minimum number of DLOs; if a folder contains fewer objects, the recommendation process will go up one level in the folder hierarchy and merge the contents of all folders below.

As mentioned before, the recommender service should also support the formation of groups, either recommending an existing community to a user or by suggesting two or more users to form a new group based on their common interests. Again the folder hierarchy should be considered for this task, since two users may work together even if they do not share all of their interests.

A major challenge in the development of the recommender system is the development of a scalable, distributed architecture. In order to allow for efficient interaction of users with their folders, the groupware services will be distributed such that a user has her private folder stored in a service located at her own institute, and group folders typically will be located at one of their members' institutes. On the other hand, the recommendation service should support a global view, in order to match users or groups located at different groupware services. Storing all relevant information in a single centralized database may not scale with a growing number of users. Thus, partitioning the database in an appropriate way will be necessary.