

Daffodil: Distributed Agents for User-Friendly Access of Digital Libraries

Norbert Gövert, Norbert Fuhr, and Claus-Peter Klas

University of Dortmund {goevert, fuhr, klas}@ls6.cs.uni-dortmund.de

1 Introduction

The Internet makes searching for literature in Digital Libraries (DLs) feasible. However, often a user has to contact several DLs to satisfy a given information need. This leads to usability problems due to the heterogeneity of the DLs. One aspect is that the information structures of the systems differ. In fact, relevant information may be spread across several DLs. The other aspect of heterogeneity is differing browsing and searching functionality, of course presented to the user through different user interfaces and query languages.

We demonstrate *Daffodil*, a system of distributed agents for user-friendly access to Digital Libraries, which overcomes these problems. Daffodil integrates several DLs freely accessible through the Internet to present to the user a single virtual DL. In addition to unifying access points and user interfaces Daffodil also produces synergies while exploiting different DLs: information from different DLs about the same document in different DLs is merged. Further Daffodil combines the functions of the underlying DLs in order to provide *high-level* search and browsing functions. As a systematization of Daffodil's functionality we use Bates characterization of information search activities on four levels of abstraction [2]:

Moves are understood as basic units in the information search process. They correspond to the commands of information systems. Examples are adding a term or a condition to a given query or following a link.

Tactics comprise one or more moves with the purpose of improving or speeding a search. One tactic is query refinement, e. g. generalizing or specializing a search term to improve recall or precision, respectively.

Stratagems exploit information structures of a given search domain, e. g. in a journal run (browsing a journal which is central to a given information need), in a citation search (following links in a relevant document), or in an author search (searching for other publications of an author, who occurred earlier in one or more relevant documents).

Strategies are plans which comprise moves, tactics, and stratagems, in order to process a complete information search, i. e. cover an information need. A simple strategy for literature search in a novel area would be an area scan in a first step, followed by an author search with authors of relevant documents found in the first step.

Most DLs provide functions on the level of moves only. From that level we can distinguish the other levels by their orientation towards a predefined target (i. e. solving a specific information search task) as well as by their *strategic* element: Tactics, stratagems, and strategies describe *plans* by which their respective target can be reached. Daffodil offers search and browsing functions on a strategic level in order to guide users

through the information structures. The following section describes the architecture of Daffodil, which implements such search activities. In section 3 we give a brief sketch of Daffodil's user interface.

2 Layered architecture of Daffodil

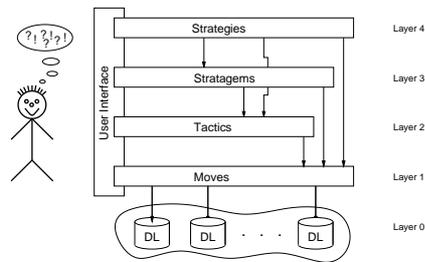


Fig. 1. Architecture of Daffodil

Figure 1 illustrates the layered architecture of our system. The bottom layer is made up of the digital libraries connected to Daffodil, each of them with its own interface. The layers one to four are the *moves*, *tactics*, *stratagems*, and *strategies* layers, each providing the respective class of functions to the user via the user interface.

For implementing the layered architecture, we developed a system of interacting distributed agents. The flexibility and extensibility gained from using an agent-based architecture enables us for example to integrate new DLs and new search services. Agent characteristics like *ability to communicate*, *adaptiveness*, and *autonomy* [4] allow for the design of such a system.

Moves

The moves layer deals with heterogeneity on a syntactical level. It provides for a uniform interface to the DLs of layer 0. For this purpose a *wrapper agent* has been implemented for each DL connected to Daffodil. The main task of a wrapper is to map the functions and information structures of the associated digital library into the system of interacting agents. Twelve DLs have been wrapped into the system so far, each with its different emphasis. For example we use CiteSeer¹ for citation searches and as a full-text service, as well as the NCSTRL² service; the ACM DL³ can be used for journal and conference runs; any of the DLs connected to Daffodil can be used for performing the standard task in Information Retrieval, i. e. searching for documents by specifying one or more query conditions w. r. t. document content and associated attributes.

Tactics

The tactics layer deals with heterogeneity on a semantical level. It produces synergies in that it combines the different classes of information objects of the DLs. Following the ideas of *multi-level hypertext* [1], we can arrange them on four levels of abstraction [3]: On the *schema* level, different attributes of documents and their respective metadata are modeled. Examples for standard schemas are the *Dublin Core* or bibliographic schemes like e. g. MARC. In Daffodil we use a slightly extended version of the BIBTEX attribute set. The *attribute value* level comprises the domain of a given attribute, e. g. a list of author names. Often such attribute domains are structured themselves, e. g. in thesauruses or classification schemes. A metadata record on the *metadata* level subsumes the attribute values of a given document w. r. t. a given schema, e. g. bibliographic attributes

¹ <http://www.csindex.com/>

² <http://www.ncstrl.org/>

³ <http://www.acm.org/dl/>

like author, title, and publication year. Finally, the *full-text* level contains the documents themselves, usually in formatted form, such as PDF or PostScript.

Search and browsing functions offered by DLs can now be identified by means of intra-level links and links between different levels. For example, an *attribute link* connects an attribute value (e. g. an author name) with documents and/or metadata records to which the value is assigned; so users may first browse the list of attribute values (e. g. the author index) and then view publications with a selected value. An example for intra-level links are *similarity links*. They exist between similar objects, either attribute values (e. g. *see also*-terms from a thesaurus) or documents; a user may look for documents which are similar to a known relevant one.

DLs can now be characterized by the levels on which information objects are provided and by the links they offer on or between these levels. Daffodil's tactics layer extends the functions of the connected DLs in that it combines the information structures of several DLs and the respective (intra-level) links into one virtual Digital Library.

Stratagems

As described in the previous sections, DLs provide a large variety of basic search and browsing functions, depending on the types of information objects they provide. They require that users are supported in using these functions effectively.

With the layers one and two we get a uniform interface for accessing the functions and information objects of the DLs. Combining them in a single system, we now can offer additional and higher-level search and browsing functions, thus providing a user-friendly access to the information structures. The stratagem layer aims at guiding the users through information structures in a given search domain. On this layer we implemented agents for seven stratagems, under them *author search* (searches for publications by focusing on authorship; users can browse authors, search for co-authors and extract authors from a set of metadata records in order to formulate new queries), *area scan* (which supports navigation through a classification scheme and provides links to documents belonging to each category), and *aggregation browse* (which focuses on tables of content of journals and conferences).

Strategies

Since strategies are highly dependent from the situation in which a user performs an information search, it is difficult to provide for a fixed set of strategies. Therefore we spent little effort here and only provided the *presearch interview* strategy which is applicable in the most general situation. The user is asked for known facts about the information objects to be searched, e. g. whether specific authors, journals, or conferences are known which might be helpful for the information search. Depending on the answers, Daffodil proposes one or more stratagems to process the user's information search task.

3 The user interface

The functions implemented in the layered architecture of Daffodil are offered to the users through the user interface. In our system the user interface is implemented by using HTML with forms and frames, thus it can be easily accessed via a Web browser. The task of the user interface is twofold: on the one hand it must present the overall functionality of the agent system to the user; on the other hand it has to visualize the results adduced by the system.

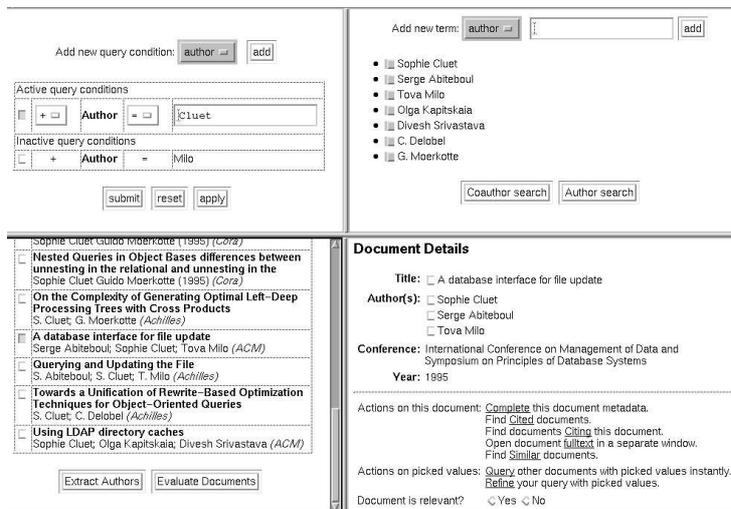


Fig. 2. User interface visualizing an author search

list of metadata records. Selecting a record displays additional information about that document in the lower right frame. The upper right frame shows the domain of a single attribute, in this case a list of author names, and allows for filtering and browsing the attribute values. Each of the frames comprise a set of functions, which can be performed on the information objects displayed, i. e. these functions implement intra-level links and links between the different levels of the multi-level hypertext.

4 Summary

We present Daffodil, which implements a virtual Digital Library, comprising information structures and functions of a (extensible) set of Digital Libraries freely accessible through the Internet. Daffodil provides for user-friendly access to contents and functions of Digital Libraries in that it offers high-level search and browsing functions to the users. Such functions guide the user through the information structures in a single-minded manner.

References

- [1] M. Agosti, G. Gradenigo, and P.G. Marchetti. Architecture and functions for a conceptual interface to very large online bibliographic collections. In *Proceedings RIAO 91*, pages 2–24, Paris, France, 1991. Centre de Hautes Etudes Internationales d’Informatique Documentaire (CID).
- [2] M. J. Bates. Where should the person stop and the information search interface start? *Information Processing and Management*, 26(5):575–591, 1990.
- [3] N. Fuhr. Information retrieval in digitalen bibliotheken. In *21. DGI-Online-Tagung – Aufbruch ins Wissensmanagement.*, Frankfurt, 1999. DGI.
- [4] Hyacinth S. Nwana. Software agents: An overview. *The Knowledge Engineering Review*, 11(3), 1996.

Figure 2 shows an author search. The screen is split into four main frames. The upper left frame presents the current query, consisting of a set of conditions. The other three main frames represent one level of the multi-level hypertext each. The lower right frame shows the query result, a