

AQUA: Query Visualization for the NCSTRL Digital Library

László Kovács, András Micsik, Balázs Pataki
MTA SZTAKI

Computer and Automation Research Institute
of the Hungarian Academy of Sciences

Department of Distributed Systems

H-1111 Budapest XI. Lágymányosi u. 11. Hungary
{laszlo.kovacs, micsik, pataki}@sztaki.hu

Abstract

AQUA is a prototype implementation of a general user interface paradigm which is capable of modelling iterative query refinement and embedding several common and advanced query techniques. AQUA has been applied to the NCSTRL (Networked Computer Science Technical Reports Library) distributed digital library.

1 Introduction

There are numerous techniques for querying digital collections ranging from searching and browsing, to relevance feedback, result ranking and so on. These techniques are usually combined, and users construct several queries until they get the desired results or the desired query expression. Recent analysis of digital library usage [2] showed that consecutive queries in a query session very often share some words, while other words are replaced. Therefore it is very important to provide a query user interface that can visualize and ease the process of building a query expression in a user-friendly way.

The AQUA project [5] at MTA SZTAKI aims at building such an interface. The first prototype of this interface has been adapted to a digital library of technical reports. The underlying software is in both cases the Dienst distributed digital library system [3]. Dienst is accessible through any standard WWW browser, and offers searching, browsing and downloading functionality in a transparent way for distributed document repositories.

2 Query Visualization

The exploration process in the digital library is broken down into a chain of operations, where each operation takes the result of the previous operation as a set of documents, and constructs a new set of documents as a result. In general, an operation can represent any search technique from boolean search to document clustering. This series of operations and results will be called the query chain throughout this paper.

For the visualization of query chains, AQUA uses a series of panels. Each panel represents an operation used to generate the next result. The first panel applies the first operation to the whole document collection. In the panel, users can choose from various simple query methods, and they can construct a query. For example, if they choose boolean search, a text box appears to enter the keywords. If they choose selection by publisher, then the list of publishers appears in the panel.

The result of the query can be represented in the panel of the query itself (e.g. search returned 238 results), and also in the next panel (e.g. browse by authors shows the authors of the previous result), as the next panel takes this result as its input for the next query. In this way, panels show not only the parts of the query expression, but also give some feedback about the intermediate results. At the end of the panel list, the final panel is a bit irregular: it has no associated query, it merely displays a selected document.

Panels are placed one after the other in a scrollable area, so the user can easily overview the whole query, and change it at any point. Typically, the query chain is grown first by adding newer and newer panels at the end. After each query users have some feedback about the result, and if those are unsatisfactory they may change the last panel or they may backtrack to change previous panels. If they change a query in the middle, then the subsequent queries are automatically recalculated, if possible, till the end of the panel list.

Constructed query chains can be saved and later reapplied. Even piping query chains one after the other

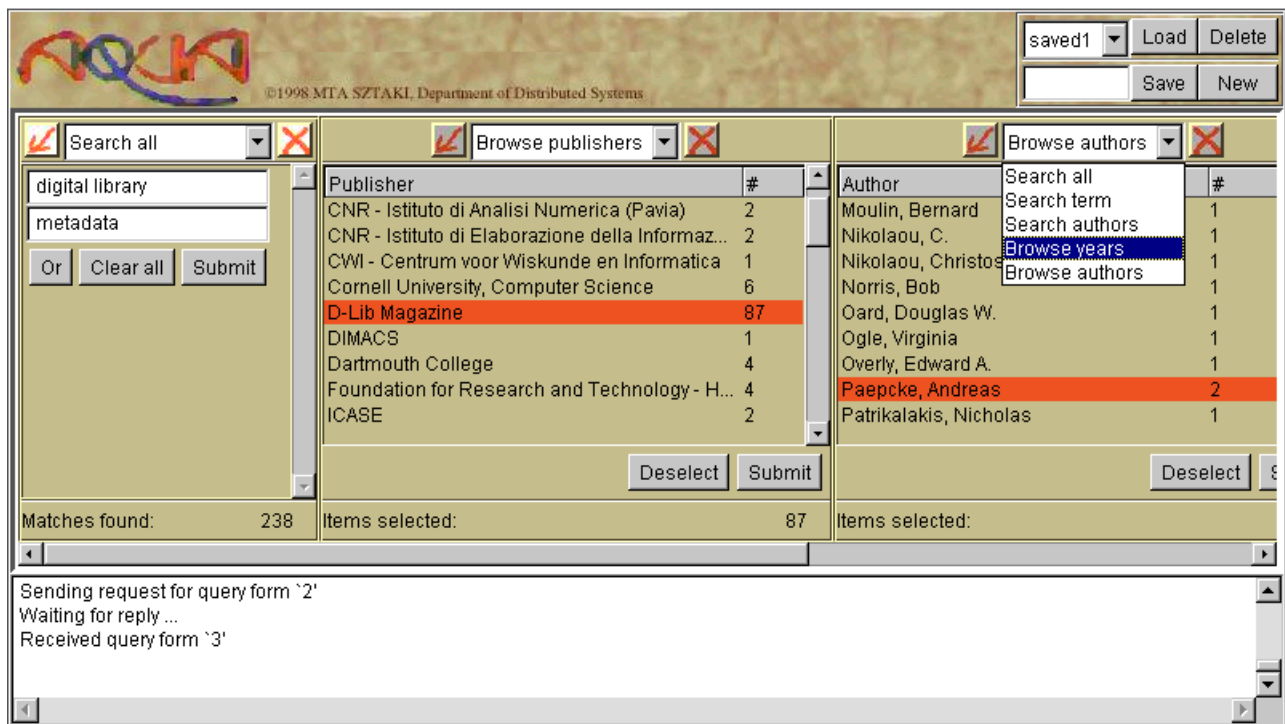


Figure 1: A query session using AQUA

might be useful. Generally, panels can support not only simple search techniques like browsing or searching, but advanced methods like clustering, query expansion, or relevance feedback. The panel area is suitable for making selections in visualized properties of a document set, for example in a graphical view of document clusters, or in some kind of visualization of keywords. Finally, in this approach different exploration techniques are not separated as in other common digital library interfaces.

In the example in Fig. 1 the first panel contains a boolean search for terms "digital library" or "metadata" with a result of 238 documents. The second panel refines the first query by selecting documents published by D-Lib Magazine. The third panel lists authors of the remained 87 documents. In the upper right corner of the AQUA window the user can save the current query chain or load a previously saved one. When a query chain is loaded each operation in the query chain is performed automatically, and each panel is updated again with the actual results.

3 Implementation

The prototype of AQUA currently works with the ETRDL [1] and NCSTRL [4] technical report collections, served by the Dienst system. Both the AQUA client and the server was implemented in Java. The client is an applet which runs in the window of a Java-enabled Web-browser, and shows the actual state

of the query chain, computed by the AQUA server. The AQUA server handles multiple client sessions, and it can access several document collections in parallel. Wrappers are used to connect digital libraries with the server, so different digital library systems can be supported with the same server.

References

- [1] S. Biagioni, J. Borbinha, R. Ferber, P. Hansen, S. Kapidakis, L. Kovács, F. Roos, A. M. Vercoustre: "The ERCIM Technical Reference Digital Library", ECDL'98, Heraklion, Greece, September 1998, LNCS 1513, Springer, 1998
- [2] Steve Jones, Sally Jo Cunningham, Rodger McNab: "An Analysis of Usage of a Digital Library", ECDL'98, Heraklion, Greece, September 1998, LNCS 1513, Springer, 1998
- [3] C. Lagoze, J. R. Davis: "Dienst: an Architecture for Distributed Document Libraries", Communications of the ACM, 38 (4) April 1995
- [4] Networked Computer Science Technical Reports Library (NCSTRL), URL: <http://www.ncstrl.org>
- [5] AQUA project homepage, <http://www.sztaki.hu/sztaki/aszi/dsd/aqua>