Wide Area Technical Report Server

Kurt J. Maly
Edward A. Fox
James C. French
Alan L. Selman

Contents

I. Introduction

Shortly before Snowbird '92 a simple request for information on electronic publication of computer science technical reports generated an enormous amount of net traffic. It appears that there are at least thirty different efforts going on, and at least as many departments who would like to have electronic dissemination but as yet do not. With an increasing number of high quality reports being produced at more and more research organizations, coupled with decreases in department operating budgets, it is becoming increasingly difficult for departments to keep their faculty informed about research at other universities. Charging for reports would merely escalate costs across the research community. Thus, a system that would enable researchers to have access from their workstations to the technical reports that are available within their own department and elsewhere has obvious appeal. In the larger sphere of activity, a number of organizations (e.g. text book publishers, computer societies) are studying new technologies for electronic publishing and multimedia access to information as well as the copyright protection issues that arise from these technologies.

Given the large interest, a workshop was held at Snowbird '92 with primary charge of developing a recommendation as to what, if anything, the computing research community as a group should do. We quickly realized that having thirty or so different systems in use is not necessarily better than having no solution at all. In short, we decided to implement a solution with existing technology in which all computer science departments who are connected to the Internet could participate. While the system we propose is in use, research groups and interested organizations will develop more sophisticated solutions to the long term problem. Specifically, we expect to have a prototype system available by Summer 1993 for a beta test involving about ten to fifteen universities.

The system we are building will consist of a central index to be kept on dual servers which will store relevant information in an easily accessible format on all technical reports for all participating departments. Each participating site will locally store its own technical reports (e.g. in Postscript, DVI or ASCII form), and will receive a copy of the software which will handle index interactions between the site and the central index facility. Faculty will not need to be involved in index maintenance.

Any user with access to the Internet can access the index facility through an interactive interface with queries about: university, author, date, CR category and keywords. For each match, selected from a ranked list, the interface will display a record containing title, authors, university, date, CR category, keywords and technical report number. The local interface will display (or store or locally print) the selected report after applying the appropriate filtering or formatting software.

To participate in providing technical reports to the system, a site needs to acquire and install the software which we will distribute, or use some equivalent compatible software. Site administrators need to require all interested local researchers to submit to the department their technical reports, after the departmental approval process, with citation details n an abstract. The department will use our software or equivalent to send citation data to the server sites, for inclusion in the central index. If the department wishes to allow access to the full–text (and diagram etc.) version of their reports, they will store the needed files locally. In that case, they will also send authentication informations and pointers to local files to the server sites, so that other sites with permission can access the full reports electronically. The system is designed to be easy to use — the submittal process can be handled by a trusted secretary.

Obviously, not all sites will capture all technical reports, and some may not be complete because certain charts and photographs may not be included. However, most documents currently being produced can easily be converted into Postscript or one of the other formats that will be supported initially, and support for other formats (e.g. SGML, rich text format) can be added later. Even if initially we capture only 60% of the available technical reports we will be better off than we are now. After one year of startup we would expect to obtain participation of about 100 sites with about 3,000 – 5,000 technical reports.

The current prototype implementation, Wide Area Technical Reports (WATER), uses WAIS as the basis for the index server (and XWAIS and SWAIS as the basis for the clients connecting to the server). It seems to be easy to use (i.e. is practically self explanatory), and can handle queries as described above and the display of reports.

II. Design

Our current design involves functions for

1) end–user access and
2) database support

Thus we separate the querying of the master index and the reviewing of documents from the creation and maintenance of technical report depositories and the master index. The philosophy for the first part is to: require no special software, rely only on commonly available software, and make it extremely simple and self explanatory. This is very important because it potentially applies to all computer science users of the Internet and must work for any conceivable hardware and software configuration. The second part applies only to a limited number of sites (on the order of a few hundred) where we can expect to install and maintain simple software. Figures 1 and 2 describe these two activities and indicate who provides what software.

Queries and Document Review

A user at any site calls the appropriate local WAIS client (e.g. XWAIS, SWAIS or PCWAIS) with queries about some technical reports. The client interacts with the master server at the central site (ODU or VPI & SU) which searches a master index and provides the client (and the user) with a list of matches. When the user indicates interesting a a particular match the report is obtained from the appropriate remote site and given to the user's WAIS client which will invoke the appropriate previewer.

Technical Report Depositories and Master Index Creation and Maintenance

The department's administrator accesses the local technical report handling system for additions, deletions, and modifications to the local structure. We will provide a copy of this software to the departments who do not have such a system but emphasize that it is not necessary for the query and reviewing part of our system. This same system will also allow local researchers to access and preview local technical reports. The part we require each participating site to have is the local index server which will transmit the local index to the master index server in a standard format (e.g. refer format plus full pathnames and representation details of files).

Approach

So far we have implemented a prototype which can handle the query part. We also have a prototype system for the local technical report handler. The overall goal of our project is to have a simple, usable and robust prototype system in place by Fall 1993. With additional support, we expect over the next three years to implement a fully developed, evaluated and operational system with the right features and with thousands of users. To this end we are preparing a proper three year proposal to NSF. In the meantime we are asking with this proposal to obtain a six month grant to develop a full prototype of the total system implemented at the four participating sites (ODU, VPI & SU, UVA, SUNY Buffalo) ready for beta test.

Finally, we wish to begin user testing of the proposed prototype system and its features and begin evaluation of the performance and scalability of the system. The latter will include such measures as: user response time, network loading, and maximum number of concurrent users.

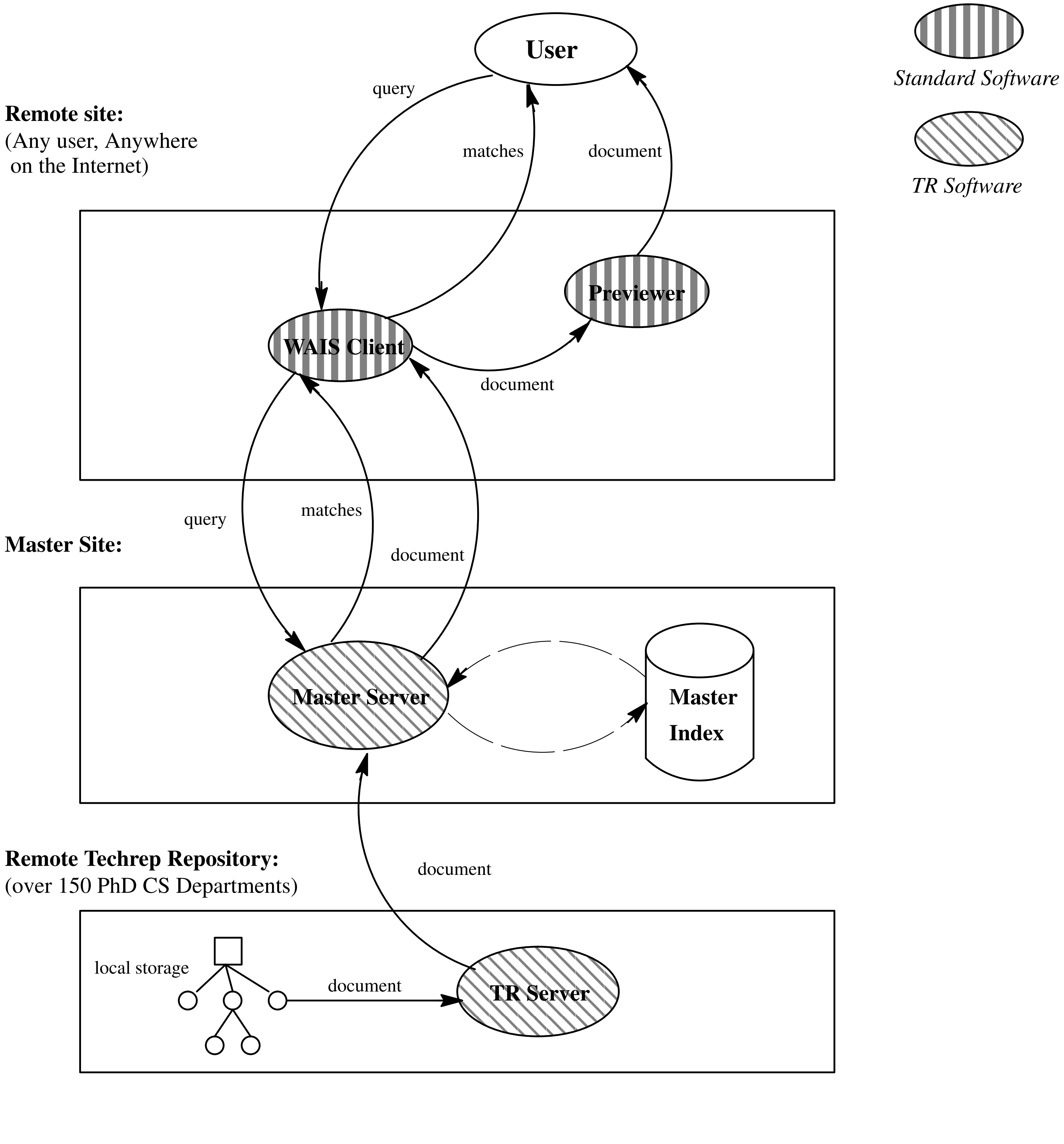

Figure 1: WATER Query and Document Retrieval Functional Diagram

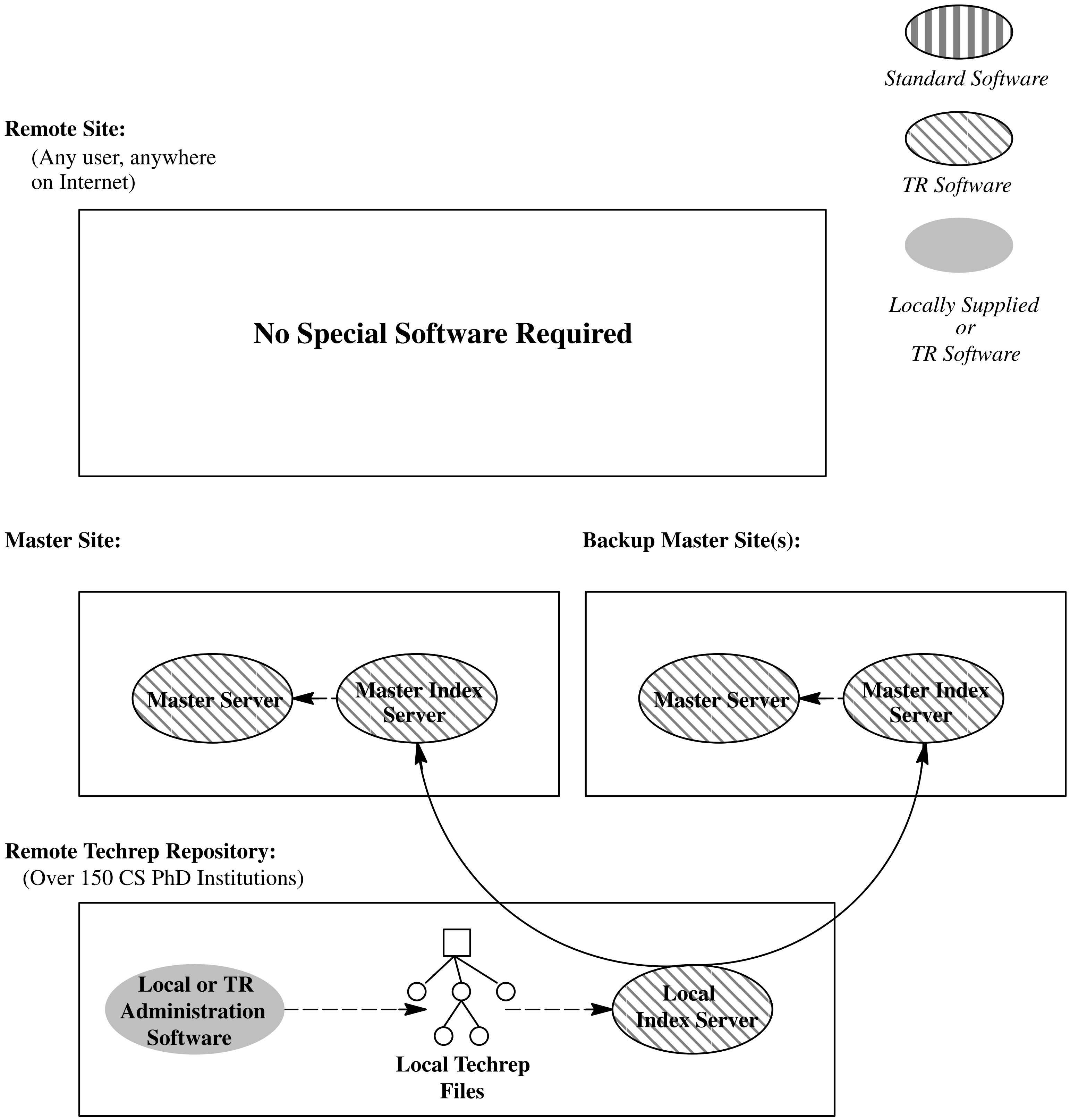

Figure 2: WATER Distributed Indexing Functional Diagram