Computer Science Department
University of Waikato, New Zealand

NZDL: The New Zealand Digital Library

PRINCIPAL RESEARCHERS

Ian H. Witten
Mark D. Apperley
David Bainbridge
Sally Jo Cunningham
Steve Jones
Lloyd A. Smith
John R. Venable

CONTACT

Ian Witten
Phone: +64-7-838 4246
Fax: +64-7-838 4155
ihw@cs.waikato.ac.nz

INTERNET HOME PAGE

http://www.cs.waikato.ac.nz/~nzdl

Funded by the New Zealand Foundation for Research, Science and Technology and the New Zealand Lotteries Grants Board.

The migration of information from paper to electronic media promises to change the whole nature of research, and in particular the methods by which people locate information. The goal of this project is to explore the potential of internet-based digital libraries. Our vision is to develop systems that automatically impose structure on fundamentally anarchic, uncatalogued, distributed repositories of information, thereby providing users with effective tools to locate the information they need and peruse it conveniently and comfortably.

WHAT IS THE NZDL?

The NZDL is a publicly-accessible system that provides full-text indexes to several substantial collections of information. The most prominent is the Computer Science Technical Report library, containing over 25,000 research reports from around 280 sites worldwide—over half a million pages of information. There are two smaller collections of English literature (500 or so books), and a collection of "frequently asked questions"—questions, with answers, that people have accumulated on thousands of topics, from classical guitar playing to sea kayaking.

HOW CAN I ACCESS IT?

The Library is accessible from our Internet home page. You can locate particular documents by specifying any combination of words in them—so-called "full-text retrieval." Search terms can be combined with the logical operators "and", "or" and "not" to create a more specific query. Alternatively a general query can be issued and the top few documents—the ones matching most closely—are returned. In either case response is very quick, and the plain text of each matching document, along with facsimile images of the first couple of pages, is instantly made available for browsing, along with a pointer that allows the original formatted document to be downloaded from its home site—which is often overseas.

HOW DOES IT WORK?

The full-text index is provided by MG, a system for compressing and indexing large collections of text (described in *Managing Gigabytes*, a book by Ian Witten, Alistair Moffat and Tim Bell, Van Nostrand, 1994). The Computer Science collection contains text extracted from PostScript documents using novel software that we have developed. To create the initial collection a computer in North America was used to download and process documents, compressing the result for economical transmission to New Zealand.

CAN I MAKE MY OWN COLLECTION?

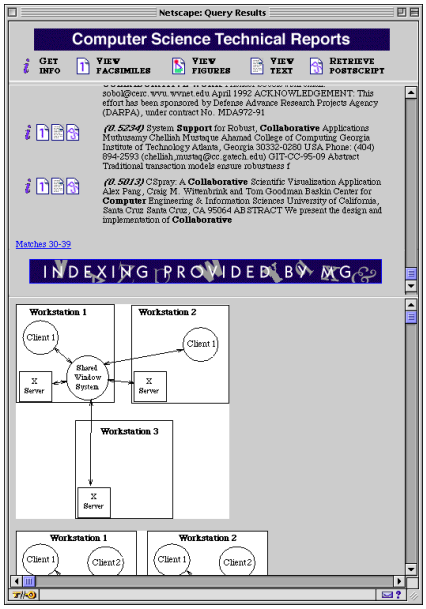
The New Zealand Digital Library software is freely available, but is still under development and has not yet been packaged in a way that makes it easily portable. It runs on Unix systems, utilises MG, and requires Perl 5.002.

Where resources permit, we are prepared to make small sample collections—up to say 500 Mbyte—available, if you can supply the raw information to us in a suitable form.

WHAT RESEARCH IS INVOLVED?

We have an extensive research programme in digital libraries. Our present work is aimed at

- developing technology for creating and automatically maintaining collections;
- monitoring usage to study library users' needs;
- looking at novel interfaces that cater to a wide spectrum of users;
- developing methods for abstracting layout and bibliographic information from document files;
- using this information to enhance presentation and for bibliometric research;
- assessing potential subject areas for public-domain collections;
- surveying and critiquing other digital library projects.



Querying the Technical Report collection: documents found (top) and figures from one report (bottom)