

# Greenstone Digital Library Software: Current Research

David Bainbridge and Ian H. Witten

Department of Computer Science  
University of Waikato  
Hamilton, New Zealand

{davidb,ihw}@cs.waikato.ac.nz

## ABSTRACT

The Greenstone digital library software ([www.greenstone.org](http://www.greenstone.org)) provides a flexible way of organizing information and publishing it on the Internet or removable media such as CD-ROM. Its aim is to empower users, particularly in universities, libraries and other public service institutions, to build their own digital libraries. It is open-source software, issued under the terms of the GNU General Public License. It is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO.

This demonstration showcases Greenstone and our current research directions—many of which are the subject of papers submitted to JCDL2004. It includes the following.

**The Greenstone Librarian Interface** is a tool that supports a variety of workflows to govern the formation, design, enrichment and maintenance of DL collections. The design has benefited greatly from experience with designing and using related tools and is now an integral component of the Greenstone software. Conference delegates are invited to gather content and use the tool to build their own collections on the spot. [Described at JCDL2003]

**Training trainers.** Training is a serious barrier to the adoption of advanced information systems in public institutions—particularly in poorer parts of the world where commercial courses are infeasible. UNESCO's Communication and Information program mounts workshops that promote the development and sharing of DL collections using Greenstone. Workshop attendees are supplied with a full set of teaching material, in printed form and also on CD-ROM, to use for local courses back home. Delegates can peruse both the online and printed material. [Short paper submission]

**Greenstone3** is a radical redesign of Greenstone, informed by our end users' and collection developers' experience over the last decade. It provides flexible ways of dynamically configuring the run-time system and adding new services. It makes it even easier for collection developers to organize and present their content. It modularizes the internal structure and simplifies the addition of new modules. The design is based on widely-accepted standards, current software practices,

cross-platform development strategies, and contemporary schemes for software modularization and dynamic updates. Delegates can test-drive a selection of sample collections. [Full paper submission]

**Realistic books** could significantly enhance DLs. A 3D visualization can give readers a better sense of holding a physical book, and provide natural ways of handling it. In real documents, certain kinds of metadata (age, usage) are manifested as visual cues. This project emulates such features to see whether they can enhance the reader's experience. Delegates are invited to interact with the book visualization. [Full paper submission]

**Collaging.** Android Kerne's "collaging" technique displays an interactive montage of images, typically sourced from the web. We have reimplemented this in Greenstone, based on images culled from documents in a DL collection, to produce a compelling, relaxing, serendipitous form of browsing.

Actual content, and actual metadata, are the driving forces behind the Greenstone project. Greenstone develops in response to real demands that stretch its configurability and extensibility. Examples range from bibliographic based collections such as the BBC Archives to full-text indexed ones such as Ulukau's Hawaiian Historic Newspapers project and the UN Humanitarian Aid collections; from audio with textual summaries such as the Hamilton Public Library's oral history collection to images and video collections augmented with textual metadata. This demonstration presents some recent collections.

**Ebooks archive.** A prototype collection of ebooks, mostly novels, from initiatives such as the Million Book Project and Project Gutenberg and stored at the Internet Archive. Mixture of formats used including DjVu files (original image and OCR'd text), XML records of metadata and plain text.

**First aid in pictures.** An exploratory collection, inspired by an instructional manual on first aid that uses no text, that investigates the feasibility of DL content for illiterate users. [Short paper submission]

**Historic maps.** A collection of historic maps of New Zealand, augmented with (amongst other things) a gazetteer of place names and geographical features. The geographic functionality is implemented as an auxiliary service using Greenstone3.