Chapter 1.10
The Greenstone Digital Library Software

Ian H. Witten
University of Waikato, New Zealand

David Bainbridge
University of Waikato, New Zealand

ABSTRACT

This chapter describes the evolution of the Greenstone digital library project through its first 10 years of development. It provides an overview of the software, which includes both production and research versions, as well as a chronological account of notable events that occurred in this period. The chapter also focuses on the tension that occurs between trying to support two versions of the software, and our strategy for resolving this conflict, that is, of reconciling production values with a research framework.

DOI: 10.4018/978-1-60960-472-1.ch110

INTRODUCTION

At the time of writing (December 2007) Greenstone—a versatile open source multilingual digital library environment with over a decade of pedigree—has a user base hailing from over 70 countries, is downloaded 4,500 times a month, runs on all popular operating systems (even the iPod!), and has a Web interface in over 50 languages. It is also a successful framework for research, as evidenced through the group's publication record.[1] How did this software project and the research team behind it reach this point? Team members often give anecdotal stories at conferences and workshops about life behind the scenes; this

chapter provides a more definitive and coherent account of the project.

This chapter is divided into three parts. First we present an overview of the software in its current form. It comes in two flavors: a production system and a research framework. Next we give a chronological account of its development from the very beginning, including origins, early adopters, and our approach to the key issues of sustainability, support, and interoperability. The existence of two flavors creates a tension—the versions compete for the resources that we can commit to their development—and the chapter culminates in a discussion of our strategy for reconciling this conflict.

Production Version

The production version of Greenstone provides the ability to create collections of digital content, display the content in a Web browser (either over the intranet or standalone from CDROM or similar), and access and search the collections that have been built. Its development was informed by the experience of non-governmental organizations (NGOs) such as Human Info and United Nations agencies such as UNESCO and Food and Agriculture Organization (FAO) in facilitating the dissemination of humanitarian information, particularly in developing countries. Through UNESCO sponsorship the software is fully documented—right down to configuration options embedded in scripts—in all six official UNESCO languages of English, French, Spanish, Russian, Chinese, and Arabic. The Web interface has been translated into over 50 other languages through the efforts of a large cohort of volunteers.

Countless digital libraries have been built with Greenstone. They range from historic newspapers to books on humanitarian aid, from eclectic multimedia content on pop-artists to curated first editions of works by Chopin, and from scientific institutional repositories to personal collections of photos and other documents. All kinds of topics are covered: the black abolitionist movement, bridge construction, flora and fauna, the history of the Indian working class, medical artwork, and shipping statistics are just a random selection.

Greenstone accommodates a wide variety of data types. Document formats include HTML, PDF, OpenOffice, Word, PowerPoint, and Excel. Metadata formats include MARC, Refer, Dublin Core, learning object metadata (LOM), and BibTeX. Greenstone also accepts a wide variety of image, audio, and video formats. Full-text indexing of all document text and metadata is the standard and supplied by default. Greenstone supports numerous interoperability standards, including open archives initiative protocol for metadata harvesting (OAI-PMH), Z39.50, and metadata encoding and transmission standard (METS). Collections can be exported to Fedora, DSpace, and MARC. See our Web site www.greenstone.org for more details.

End users experience Greenstone through a Web interface such as the one shown in Figure 1, taken from the Human Info NGO's *Humanity Development Library*. Documents in this collection can be searched by chapter title, as well as the standard full text search by chapter or book. Alternatively, users might choose to browse alphabetically by title or hierarchically by subject or organization. In Figure 1(a) the user has searched within chapters for the word "environment" and a ranked listed of matches has been displayed. In Figure 1(b) they are viewing the document that results from selecting the second matching item, that is, Chapter 3 of *Teaching Conservation in Developing Nations*. Greenstone collections are designed individually, depending on the document formats and metadata available and the facilities that the designer wants to offer the user. Figure 1 shows a representative example, but other collections often look completely different.

Figure 2 shows the Greenstone librarian interface (GLI), an interactive application for creating and maintaining collections such as the Humanity Development Library. Through a system of