

Research Experiences for Undergraduates Supplement

Portable Biodiversity Information Server

(a) Overview

In a previous REU grant (Heidorn, 2000) a group of undergraduate students developed a set of thesaurus tools for a botanical information retrieval and visualization system (Heidorn, 2001). In that project, the students built a parser to extract information from the *Categorical Glossary for the Flora of North America Project* (Kiger & Porter, 2001) and the *Online Glossary of the Flora of Australia* (McCusker, 2002). These tools are used to provide automatic and manual query expansion as well as inline definitions. From the study of different user groups including high school teachers and professional botanists, these tools have proven to be useful (Heidorn et al., in press).

There are five main components of the REU supplement. They involve enhancing the functionality of the previously developed thesaurus tools based on end user input. The first enhancement is the addition of definitions, synonyms and related terms and images that are more appropriate to the high school students and citizen scientists. This information in the previously installed thesauri is geared toward a professional audience. The second enhancement is the addition of an online glossary and thesaurus entry tool to allow teachers to add definitions and synonyms to the online thesaurus over the web. The third thesaurus enhancement is the inclusion of a new text as a source of definitions. We have copyright permission to create online versions of "Vascular Plant Systematics" (Radford, 1998), as well as the glossary of "Taxonomy of Vascular Plants" (Lawrence, 1951). The fourth thesaurus enhancement is, for performance reasons, a shift from SOAP to cgi for online delivery of selection items. The fifth modification is actually a series of changes to the thesaurus tools to operate in a mobile networking environment.

The effectiveness of this new functionality will be evaluated through the evaluation mechanisms of the main project. This will include training sessions for teachers, students and professional botanists on the use of the system, focus groups, experimentation and field observations.

(b) Nature of Student Activity

Each of the system development tasks, programming tasks and information generation tasks will be kept simple and modular so that new components can be developed every few weeks. Perl will be used to manipulate the original text of the glossary and insert it into a MySQL database. This database will be used to generate menus and perform query expansion. Java is used in the system interface to provide users with menus of search terms and other options. The students will participate in the spring testing of the thesauri. The system interface tracks all user actions. As time permits these students will help extract data from the log files for analysis.

In this project we propose to build a set of vocabulary support utilities that extend the functionality of botanical glossaries to improve access and usability of online floristic databases. We will make innovative use of electronic glossaries and thesauri from several floristic publishing efforts. A number of Flora are being made available in electronic format (Flora of North America, 1993a; Flora of China, 1994). This is motivated by the hope that electronic Flora will be more readily accessible by more people (Wilken et al., 1989). One problem is that these flora are written for scientific purposes and as such make use of scientific vocabulary that is not readily understood by non-professionals. Even botanists can differ on the selection of descriptive

terms for the same item. Botanists also develop specialized vocabularies or context dependent meanings of words in specialized groups

Given the complexity of describing natural objects, the historical use of descriptive terms, and the sometimes subtle distinctions that need to be made between species, it is difficult to impose a prescriptive controlled vocabulary or key-words as is used in indexing and information retrieval work. In addition, the same term may have different meanings in different contexts. The leaf blade of a deciduous tree is very different from the meaning of "leaf blade" when an author is describing a fern. This freedom in word selection by authors poses a double problem for people who later wish to perform full-text searches on the flora. Even a professional botanist can not know which of a set of synonyms was used in a particular treatment. A non-botanist may not even be able to think of any of the "accepted" technical botanical terms. The second problem is that there is no guarantee that when a botanist searches the text later, that they will even enter the same collection of synonyms unless the glossary is consulted and the author followed it as well.

This proposal includes five extensions of the online glossaries/thesauri to facilitate access by professional botanists and non-botanists alike. In this project we will then evaluate this new functionality to see if it helps different classes of users perform tasks that require information that resides in the flora. The main task will be species identification and selection of native and invasive plants appropriate for a biodiversity surveys in schools and Illinois collections.

(b1) Definitions and Term Categories

The first enhancement is the addition of definitions, synonyms, broader and narrower terms and images that are more appropriate to the high school students. This information in the previously installed thesauri including the Categorical Thesaurus is geared toward a professional audience. High school teachers and professional botanists have used and evaluated the previous version of the thesaurus. Both groups reported that the terminology of the taxonomic descriptions in the Flora of North America, would be useful for professional botanists and well trained teachers but the online definitions, synonyms and other semantic categories were all too specialized to be useful for either adult professionals or students. New taxonomic descriptions are being created in a related Institute for Museum and Library Services (IMLS) funded project called OpenKey (<http://www.isrl.uiuc.edu/~openkey>).

(b2) Online Definition and Thesaurus Entry

The second enhancement is the addition of an online glossary and thesaurus entry tool to allow teachers and volunteers to add definitions and synonyms to the online thesaurus over the web. Teachers frequently include terminology exercises in their classes. This interface will allow their definitions of synonymy, broader and narrower words to be include din the electronic thesaurus. This in turn can be used to facilitate automatic query expansion, as well as inline definitions for the use of these words in botanical documents. Security and verification mechanisms will be required.

(b3) New "Printed" Thesaurus Integration

The third thesaurus enhancement is the inclusion of a new text as a source of definitions. We have copyright permission to create online versions of Vascular Plant Systematics (Radford et al, 1998), as well as the glossary of Taxonomy of Vascular Plants (Lawrence, 1951). These materials will be parsed using software from prior grants (DBI-9982849 <http://www.biobrowser.org>) and included in the online glossary.

(b4) CGI-based XML Selection Lists

The fourth thesaurus enhancement is, for performance reasons, a shift from SOAP to cgi for online delivery of selection items. The procession of selection lists in the server for use on client

machines is too slow for real time interaction. The mechanism will be replaced with a more direct cgi mechanism with results returned in XML or plain English format.

(b5) Mobil Thesaurus

The fifth modification is actually a series of changes to the thesaurus tools to operate in a mobile networking environment. One of the new main objectives of a ITR grant is to make information gathering tools available in the field through wireless Ethernet and portable field server and handheld devices.

(c) Research Environment and Undergraduate Mentoring

The REU students will become part of the project development team. They will be housed in the Information Systems Research Lab. They will have access to Unix and Microsoft Windows project computers. Their duties will include the design, development and developer testing of software. All students will be included in weekly design meetings. Where possible, these students will attend focus groups and design sessions with the target user group including high school science teachers, biodiversity survey volunteers and professional botanists. Over the past two years audio and video tapes have been collected of all of these user groups conducting biodiversity surveys, retrieval system use and focus groups. The principal investigators and graduate student design team will meet to discuss implementation priorities based on these tapes. The undergraduates would be included in the design meetings.

Information Systems Research Lab (<http://www.isrl.uiuc.edu/>) is a service unit for many projects in GSLIS. The unit supports research into and about the design, impacts, analysis, and evaluation of information technologies, including information and its properties, information services and access, and the creation/management of information content. Along with focused, single-area efforts, ISRL supports and facilitates multidisciplinary and collaborative research that involves GSLIS investigators across differing subject areas, as well as GSLIS collaborations with researchers in other academic and organizational units at UIUC and elsewhere. A brief discussion of REU funding is provided in the overview.

For three years Bryan Heidorn, the PI, taught undergraduate courses in information science at the University of Pittsburgh. These included courses in Computer Architecture and Machine Language, Systems Analysis and Design and Human Information Processing. While the Graduate School of Library and Information Science is primarily a graduate school, there is a new undergraduate minor in Information Technology Studies. The PI teaches one course in the this program, The Cognitive Foundations of Information System Design.

(d) Student Recruitment and Selection

In the fall of 1999 the Graduate School of Library and Information Science began a new Undergraduate Minor in Information Studies. The program is growing quickly and students come from all departments across campus. Some of the students are taking new classes in information organization and management. It would be desirable to include some of these students in the research projects in GSLIS. We will recruit from students in the undergraduate minor first using electronic mailing lists, electronic bulletin boards and in-class announcements. Minorities, persons with disabilities and women will be encouraged to apply. We will also recruit undergraduate students who have worked on other grants with the PI. One of these students is a female minority and two others are also women.