Botany 2002 http://www.botany2002.org/

August 4-7, 2002: Annual Scientific Conference

Pyle Conference Center
University of Wisconsin
Madison, Wisconsin

Taxonomic Description Creation, Search and Display in XML

P. Bryan Heidorn, Hong Cui, Bei Yu, Jingbo Wu and Hong Zhang,
University of Illinois
Graduate School of Library and Information Science
501 East Daniel Street
Champaign, IL 61820

A research group at the University of Illinois (UIUC) is developing a set of processing and search tools to facilitate the use of taxonomic description data (http://www.biobrowser.org). This information, which has traditionally only been available in paper format in research libraries, is now be available on the World Wide Web using advanced search tools for scientists, students and the general public. We will use components of the new software system to facilitate access to collection's information at the UIUC Herbarium and the North Carolina Botanical Garden. In this project, information from multiple full-text sources is converted to a standard XML format. The sources include *The Flora of North America*, *The Categorical Glossary of the FNA (CGFNA)*, *The Trees of Illinois* and *The Butterflies of Illinois*. The resulting files are indexed with an open-source full-text indexing engine. This software allows users to find words and phrases in any subsection of the original descriptions. Searchers may place their search terms (taxonomic descriptions) into any sub-part of a query Document Type Definition (DTD). Queries placed at the top of a hierarchy describing a document search the entire document. Queries may also be placed at any more specific node. For example, users might search for the word "blue" in the description of flowers that are eaten by a butterfly but not the word "blue" in the description of the wing color. We converted the *Species Plantarum* and *CGFNA* from text format into a relational database format. This database is used to add inline definitions and to support automatic and manual query expansion. In later work, XSL will be used to convert DELTA files to XML. In a project beginning soon, we will collect data in this open-format to facilitate the interchange of taxonomic descriptions and holdings data between herbaria.