



. Introduction

ENABLE is a multidisciplinary project funded by the NSF's National Science Digital Libraries program. The major goal of this project is to apply advances in digital library technologies to the emerging domain of bioinformatics, and develop novel interaction means that support learning based on identifying and visualizing associations among key dimensions of bioinformatics resources. Recently, an ENABLE Knowledge Base prototype was developed. Features of the current prototype include automatic bioinformatics resource harvesting, basic information retrieval operations, scattergather clustering, and force-directed layout.

2. System Architecture

The ENABLE Knowledge Base consists of two major components:

- bioinformatics resource collection
- bioinformatics resource management

The resource collection component automatically collects and refines information of on-line bioinformatics education resources. Based on the collected education resources, the resource management component provides learning services to its users through a visualization interface.

Extensible Networked Association-based Bioinformatics Learning Environment

Yong Liu¹, Weimao Ke², Javed Mostafa² ¹Computer Science Department, Indiana University ²School of Library and Information Science, Indiana University



Figure 1. ENABLE Knowledge Base Architecture

Figure 1 shows the architecture of the ENABLE Knowledge Base. The part over the dashed line represents the resource management component while the part below the dashed line represents the resource collection component. Online bioinformatics education resources are collected in the form of HTML documents. In the resource collection component, a bioinformatics resource data processor parses the HTML documents automatically, extracting attributes such as title, description, and URL for each document. It then sends requests to LUCAS[1], a concept extraction web service, to extract keywords for each document. After the keyword extraction is done, the refined resource data is sent to database for later use.

The resource management component has a 3-tier architecture. The front end is a Java applet-based GUI running on a Javaenabled web browser. The intermediate tier is Java application server, running servlet to process client side requests and send the responses back to the client. The back end is an Oracle database, which holds all the resource data, executes queries issued from the servlet and sends results back. The applet running on the web browser provides rich UI capabilities to users. Currently a visualization technique called force-directed layout is introduced to visualize relations among keywords of all the resources stored in database. Besides, a browsing method called scatter-gather is implemented to help users locate their interested resources through online reclustering operations. The most compute-intensive tasks like term-term similarity matrix computation are taken care at the intermediate tier by the servlet to guarantee a light weight client side.

3. Continuing Research

The major performance bottleneck in the current ENABLE prototype is the online reclustering process of the scatter-gather browsing. A quadratic time complexity clustering algorithm is used for current online reclustering. We are now working on implementing a linear time online reclustering algorithm for the scatter-gather browsing. In addition, we plan to develop a parallel clustering algorithm to further speed up the online reclustering process.



[1] Yueyu Fu, and Javed Mostafa. Toward Information Retrieval Web Services for Digital Libraries. IEEE/ACM Joint Conference on Digital Libraries 2004, Tucson, Arizona, 2004

[2] Jeffrey Heer, Stuart Card, and James Landay. Prefuse: A Toolkit for Interactive Information Visualization. Submitted paper draft, April 2004

http://enable.slis.indiana.edu

Figure 2. ENABLE Knowledge Base GUI

Reference

Project Website