1. 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, scatter-gather 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.

The resource management component has a 3-tier architecture. The front end is a Java applet-based GUI running on a Java-enabled 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.

Reference


Project Website

http://enable.slis.indiana.edu