ScholarWorks

'An Application for Downloading and Integrating Molecular Biology Data

Login (/login)

- IUPUI ScholarWorks Repository
- →
- School of Informatics and Computing
- →
- Informatics Theses and Dissertations
- →
- Informatics Graduate Theses and PhD Dissertations
- →
- View Item

An Application for Downloading and Integrating Molecular Biology Data

Fontaine, Burr R.

Name: view.pdf
Size: 321.0Kb
Format: PDF

View/Open

Permanent Link:	http://hdl.handle.net/1805/369
Date:	2004-07
Keywords:	downloading ; molecular biology data ; integrating data

Abstract:

INTRODUCTION AND BACKGROUND Integrating large volumes of data from diverse sources is a formidable challenge for many investigators in the field of molecular biology. Developing efficient methods for accessing and integrating this data is a major focus of investigation in the field of bioinformatics. In early 2003, the Hereditary Genomics division of the department of Medical and Molecular Genetics at IUPUI recognized the need for a software application that would automate many of the manual processes that were being used to obtain data for their research. The two primary objectives for this project were: 1) an application that would provide large-scale, integrated output tables to help answer questions that frequently arose in the course of their research, and 2) a graphic user interface (GUI) that would minimize or eliminate the need for technical expertise in computer programming or database operations on the part of the end-users. In early 2003, Indiana University (IU), IBM, and the Indiana Genomics Initiative (INGEN) introduced a new resource called Centralized Life Sciences Data Services (CLSD). CLSD is a centralized data repository that provides programmatic access to biological data that is collected and integrated from multiple public, online databases. METHODS 1. an in-depth analysis was conducted to assess the department's data requirements and map these requirements to the data available at CLSD 2. CLSD incorporated new data as necessary 3. SQL was written to generate tables that would replace the targeted manual processes 4. a DB2 client was installed in Medical and Molecular Genetics to establish remote access to CLSD 5. a graphic user interface (GUI) was designed and implemented in HTML/CGI 6. a PERL program was written to accept parameters from the web input form, submit queries to CLSD, and generate HTML-based output tables 7. validation, updates, and maintenance procedures were conducted after early prototype implementation RESULTS AND CONCLUSIONS This

application resulted in a substantial increase in efficiency over the manual methods that were previously used for data collection. The application also allows research teams to update their data much more frequently. A high level of accuracy in the output tables was confirmed by a thorough validation process.

Description:

Submitted to the faculty of the University Graduate School in partial fulfillment of the requirements for the degree Master of Sciences in the School of Informatics Indiana University July 2004

This item appears in the following Collection(s)

- Informatics Graduate Theses and PhD Dissertations (/handle/1805/303)
- Informatics School Theses and Dissertations (/handle/1805/954)

Show Statistical Information (#)

My Account

- Login
- Register

Statistics

- Most Popular Items
- Statistics by Country
- Most Popular Authors

FULFILLING *the* PROMISE