

EGOMiner: A Genomics and Proteomics Data Computation and Interpretation System for Biomedical Applications

Weimin Feng¹, Gaurav Tuteja², May D. Wang^{1*}

¹Wallace H. Coulter Department of Biomedical Engineering,
Georgia Institute of Technology and Emory University

²School of Electrical and Computer Engineering, Georgia Institute of Technology
313 Ferst Drive, Atlanta, GA 30332, USA

Abstract

EGOMiner, stands for Enhanced Gene Ontology mining system, is a CORBA based tool developed for genomic and proteomic data analysis. One key function of EGOMiner is to provide biological data interpretation based on Gene Ontology with quantitative and statistical analysis and visualization by direct acyclic graph. Cross comparison of multiple experimental studies is supported. The other function is quality analysis of microarray chip images. The input is raw chip data and output is visualization of the data quality of the chip image. This system significantly improves the GoMiner system that was designed and developed by the authors.

1. Introduction

The advancement of modern genomics and proteomics technologies has produced high-throughput gene and protein data that are waiting to be analyzed and interpreted. Specifically, to study molecular interactions and pathways of biological systems, we need to integrate these heterogeneous data and interpret their functions. Our Medical Informatics and Bioimaging Lab (MIBLab) has been researching novel genomic, proteomic, and metabolomic data analysis methods and professional software systems for disease diagnosis and drug target study. We have designed and developed a novel integrated bioinformatics system, EGOMiner, to compute heterogeneous "omic" data and to find biological important functions for study. The design criteria are scalability, high performance,

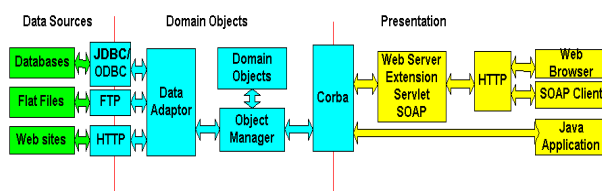


Figure 1 EGOMiner Multi-Tier System

ease of use, and interoperability with other bioinformatics systems.

2. System Design and Development

EGOMiner is designed to be a web-based system with multi-tier architecture (Figure 1). The novel architecture design makes it more flexible in system scalability and in handling performance issues. For web-based applications, large client connections can be handled based on load balancing technique. The application server, which contains domain objects, object manager and adaptors, can be scaled up from desktop computer to super computers or clusters. By server replication, the system can be further scaled up to provide more computing power and to enhance system availability.

To solve interoperability issue between different applications, EGOMiner was implemented based on CORBA standard. CORBA [2] is an industrial standard for distributed software architecture and it is platform and language independent. Applications register their COBRA objects to naming service, through which clients can get services from the objects, which are implemented by CORBA servants. The application server and clients exchange data using XML [3] format. Also, EGOMiner is

designed to have, a uniformed user interface, through web-browser, to access different applications. This releases the users from the burden of downloading or upgrading the software versions to their local machine. In each genomics or proteomics disease diagnosis or drug target study, the users can submit genes through the web browser such as Internet Explorer or Netscape. The EGOMiner system server maintained by MIBLab will analyze the data and provide biological function interpretation to the users. The upgrade of the algorithms or software system will be transparent to the users. Last, the EGOMiner system is designed to be kernel based with application plug-ins.

3. Case Study

In 2002, we have designed and developed a prominent software system, GOMiner, for our collaborators at National Cancer Institute of NIH [1]. It has been designed based on a two-tier architecture: the database server, and the application logic plus user interface. We have been deploying GoMiner since 03/2003 (<http://www.miblab.gatech.edu>). During the past months, the limitation of such a system has been shown that we have been getting requests regarding downloading the software system on a daily basis, and we also got complains on incompatibility when new version of the tool was deployed due to upgrade.

With the new software architecture, we redesigned the GoMiner application as a application plug-in for EGOMiner. Then with the diagnosis-based gene marker selection system we developed in MIBLab with clinicians at Emory University School of Medicine, omniMarker, and one proteomics data analysis system ProteiNavi developed in house, the biological and medical researchers can submit their genes or protein markers through the web interface. The genes will be processed on the backend application server, and the analysis and visualization results will be returned to the web page in a user-friendly manner. System maintenance (e.g. database and software upgrade) and performance tuning will be transparent to

users and will be performed by system administrators on server side.

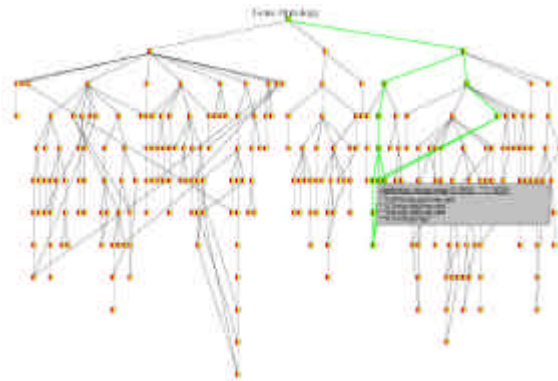


Figure2 Case Study: EGOMiner Performs Multi-Group Comparative Study

4. Future Work

This system was being designed to improve the usability of the interface. Ultimately, the goal is to extend the system to be a key test-bed for next generation bioinformatics system research, which will be platform independent and have flexibility, scalability, and extensibility to increase life span of the software system.

References

1. Zeeberg BR, **Feng W, Wang G, Wang MD**, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, and Weinstein JN, GoMiner: A Resource for Biological Interpretation of Genomic and Proteomic Data, **Genome Biology**, 4: R28, April 2003.
2. CORBA, OMG, <http://www.corba.org>
3. XML, W3C, <http://www.w3.org/XML>