Xenie: Integration of Human 'gene to function' information in human readable & machine usable way

TaeJin Ahn

http://bric1.postech.ac.kr:8989

Abstract

Xenie is the JAVA application software that integrates and represents 'gene to function' information of human gene. Xenie extracts data from several heterogeneous molecular biology databases and provides integrated information in human readable and machine usable way.

We defined 7 semantic frame classes (Gene, Transcript, Polypeptide, Protein_complex, Isotype, Functional_object, and Cell) as a common schema for storing and integrating gene to function information and relationship. Each of 7 semantic frame classes has data fields that are supposed to store biological data like gene symbol, disease information, cofactors, and inhibitors, etc.

By using these semantic classes, Xenie can show how many transcripts and polypeptide has been known and what the function of gene products is in General. In detail, Xenie provides functional information of given human gene in the fields of semantic objects that are storing integrated data from several databases (Brenda, GDB, Genecards, HGMD, HUGO, LocusLink, OMIM, PIR, and SWISS-PROT).

Although Xenie provide fully readable form of XML document for human researchers, the main goal of Xenie system is providing integrated data for other bioinformatic application softwares.

Technically, Xenie provides two kinds of output format. One is JAVA persistent object, the other is XML document, both of them have been known as the most favorite solution for data exchange. Additionally, UML designs of Xenie and DTD for 7 semantic frame classes are available for easy data binding to other bioinformatic application systems.

Hopefully, Xenie's output can provide more detailed and integrated information in several bioinformatic systems like Gene chip, 2D gel, biopathway related systems. Furthermore, through data integration, Xenie can also make a way for other bioiformatic systems to ask 'function based query' that was originally impossible to be answered because of separatly stored data in heterogeneous databases.

CURRICULUM VITAE

TaeJin Ahn

Graduate student (Master Program) Department of life science,

Postech, Pohang

Tel: +82-18-528-6040, +82-54-279-8194

email: tj@bric.postech.ac.kr

Work:

1999-present: **Postech**, Pohang.

Master degree student at Bioinformatics of Life Science Department, Postech.

1. System conceptualization, requirement, design, coding of data integration software "Xenie". Software was written for integrating general biological knowledge related to specific Human gene from several molecular biology databases. Xenie

2. Porting PIR, SWISSPROT flat file data to local DBMS(Oracle) using JAVA and JDBC technology.

3. Software requirement and Structure Design of the Unified Molecular Biology Search Engine "GeneNet". Software was written for Integrated search for several databases "GenBank", "Blocks", "KEGG", "PDB". GeneNet.

4. Building an Integrated Encyclopedia of specific cell in higher eukaryote. Building a way to integrate knowledge and data in biological domain with concept of knowledge representation. Representing knowledge of gene to function abstract for recomb2000

5. Consulting researchers from biological and medical sciences on Searching Molecular Biology Databases on the Web.

Education:

1999-present:

> **Pohang university of science and technology (POSTECH)**, Pohang. Department of Life Science. Laboratory of Prof. Hong Gil Nam.
>
> Thesis: Representing and integrating gene to function knowledge. M.Sc. course, Jan. 1999-present.

1995-1999:

> **Handong university**, Pohang. Department of bioscience and food engineering.
>
> Thesis: Software engineering in biology domain. B.Sc. degree, Feb. 1999. G.P.A 4.18/4.50 .

Awards and Honors:

1999   President's prize for excellence

Course work:

Biology: major

Organic and Analytical Chemistry | Biochemistry 1 | Biochemistry 2 | Biophysical Chemistry | Laboratory safety | Biochemical calculation | General Microbiology | Applied Microbiology | Applied microbiology lab | Laboratory in Microbiology | Molecular Biology 1 | Molecular Biology 2 | Molecular Biology lab | Protein purification | Instrumental analysis lab | Physiology | Cell Biology | Immunology | Immunology lab | Developmental Biology | Psychology | Foodchemistry | Food Processing | Food Engineering Lab | Food Analysis Lab | Sitology | Advanced biochemistry | Advanced molecular biology | Advanced cell biology | Advanced immunology | Advanced animal genetics | Proteomics and molecular network | Bioinformatics |

Computer Science: minor

Modern Mathematics Statistics | Unix & C programming | Data Structure | Algorithm Analysis | Artificial Intelligence | Object-Oriented Programming | Software Engineering | Biocomputing Using Matlab | Advanced Artificial Intelligence | Advanced Computer Algorithm | Database Systems | Objected Oriented Design Using UML -oo226

Experience and skills

Bioinformatics: Integrated Database development. Knowledge representation in biological domain. Using public databases, services and software in Molecular Biology and Genome. Consulting on usage of these resources.

Programming: SCJP(Sun Certified JAVA Programmer). Objected Oriented Designing and Development. Text manipulation, data integration, database development and network programming using JAVA. XML design, composing, parsing. HTML composing.

Biology: Biochemistry, Protein Purification, Sequencing, Enzyme kinetics.

Publications

1. "Building integrated pathway genome database in cell type specific manner", Current in Computational Molecular Biology p89, Universal Academy Press Inc. Recomb2000 abstract

References available on request.