

BioStore: A Repository System for Registering and Distributing Public Biology Databases

Hongseok Tae¹, Jeong-Min Han², Bu-Young Ahn³ and Kiejung Park^{1*}

¹Information Technology Institute, SmallSoft Co. Ltd., Daejeon 305-728, Korea, ²Korea Institute of Oriental Medicine, Daejeon 305-728, Korea, ³Korea Institute of Science and Technology Information, Daejeon 305-728, Korea

Abstract

Although abundant biology data have been accumulated in public biology databases, such as GenBank and PIR, few easy-interface services are provided for users to access or update them. We have developed a system, named BioStore, that is composed of several programs to aid users to not only access public data but also share their own data easily. The service can be used for maintaining a local database as a repository of raw data files of several public databases and distributing the data files to other users. Currently, BioStore manipulates major bio-databases and will expand to include more databases and more useful interfaces.

Availability: This system is available from <http://www.ccbb.re.kr/biostore>

Keywords: BioStore, FTP, GenBank, repository

Introduction

Biology data, especially DNA and protein sequence data, are very rapidly increasing. Most data have been submitted to major bioinformatics databases, such as GenBank (Benson *et al.*, 1994), PIR (McGarvey *et al.*, 2000), and PDB (Berman, 2000), which are distributed via FTP services to a lot of bioinformatics sites and users. As more and more bio-databases (Jeon *et al.*, 2005; Kim *et al.*, 2008) are constructed and a large amount of data files is kept in each database, most biology researchers have difficulty in searching for information and in sharing their own data with others.

CCBB (Center for Computational Biology and Bio-

informatics) of the Korea Institute of Science and Technology Information has been providing an FTP server and analysis services for major bio-databases. We have developed BioStore as a repository system by implementing a few programs to register user data, including public bio-data at CCBB. Researchers can easily register their own data with BioStore through the system to share them with others.

Features and Results

BioStore has the following features:

- bio-database mirroring
- directory browsing
- data download
- quick search
- xml viewer
- user data registering

BioStore is composed of 2 parts (Fig. 1), a data listing part and a data registering part. The data listing part lists entry names that are registered with BioStore of the CCBB server. The main web page of BioStore has a user-friendly web interface and similar features to FTP programs (Fig. 2A). The left frame shows data directo-

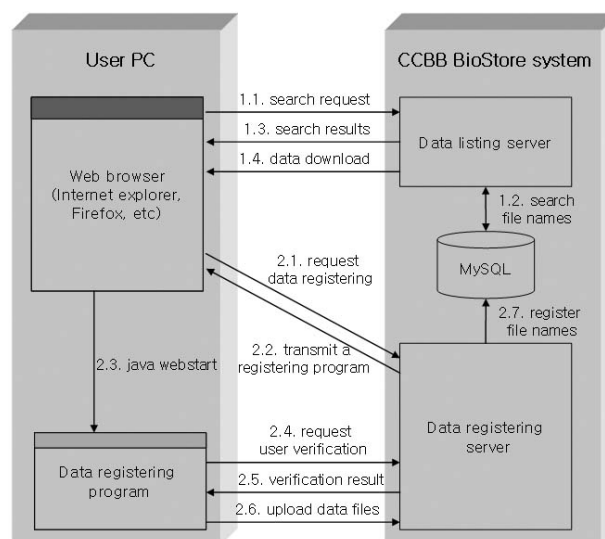


Fig. 1. The system structure of BioStore to access and to register data.

*Corresponding author: E-mail kjpark@smallsoft.co.kr
Tel +82-42-864-2524, Fax +82-42-385-9240
Accepted 20 February 2009

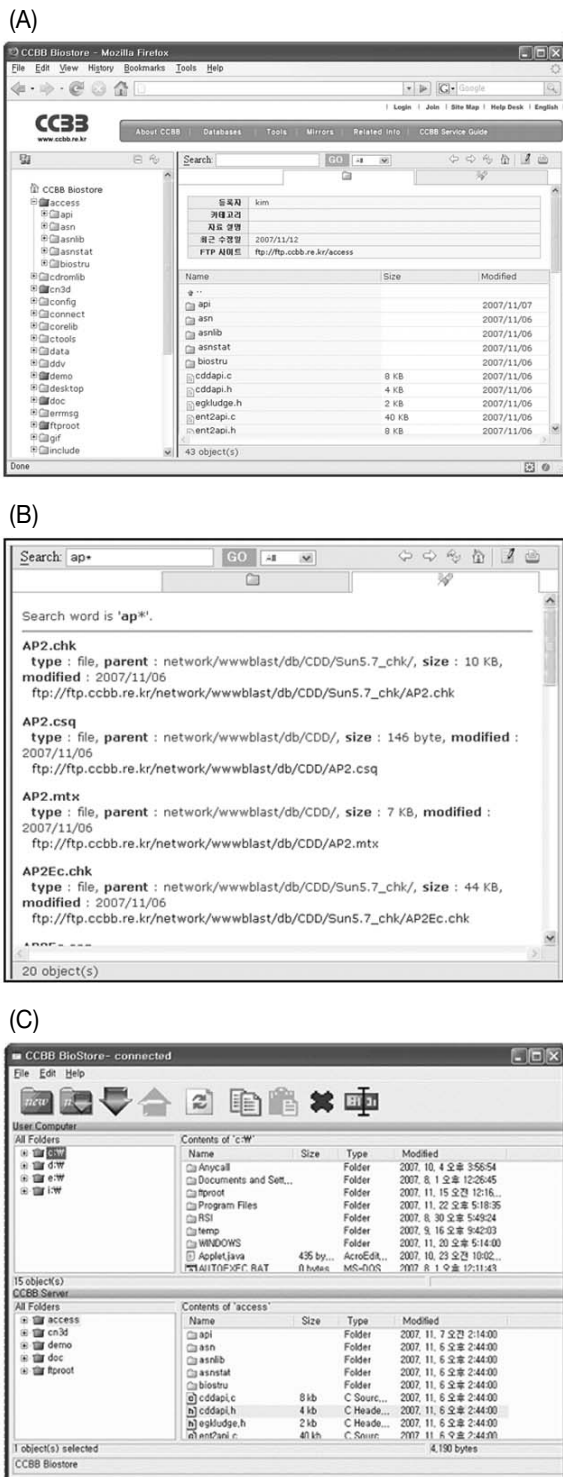


Fig. 2. (A) Main web page of BioStore. The left frame shows data directories using a tree structure, and the right frame shows the description and contents of a selected directory. (B) Search interface. Users can search data files using keywords. (C) Data-registering interface. Users can easily register their own data with BioStore.

ries on a tree structure, and the right frame shows the description and contents of a selected directory. When a user logs into the CCBB system and a directory has been registered by the user, it is colored deep blue. Because researchers have difficulty in finding data files from the FTP sites of major bio-databases due to their complex structures, BioStore provides a quick search tool with a file name search from BioStore (Fig. 2B).

The data registering part (Fig. 2C) stores the user data files to the FTP server in CCBB and registers the file names to the MySQL database. When a login user requests data registration, BioStore transmits a registration client program to the user PC, which launches automatically. The process is carried out by java web start technology. After the program verifies the user information, it connects to the data registration server of BioStore and provides upload and download interfaces. When the user uploads data files to the server, the data registration server keeps the files in the CCBB FTP server and registers data file names to the MySQL database. Then the files can be distributed by the data listing server.

Discussion

Although most bio-databases have their own FTP services to distribute data files, researchers face difficulties in finding appropriate data files due to huge amounts of data in each bio-database. BioStore has been developed to aid researchers to not only access the data files of public bio-databases but also share their own databases easily. With user-friendly interfaces, the system provides useful tools, including a search tool, an XML viewer, and a data registration feature.

The system currently provides major public bio-databases, including GenBank and PDB, and will integrate more bio-databases, including user databases, which are produced by many researchers.

With the increase of data files, the structure of directories will be more complicated, and better interfaces and features will be developed through further development.

Acknowledgements

This work was supported by MOST-through-a KSITI grant and by MEST and KOTEF through the Human Resource Training Project for Regional Innovation.

References

Benson, D.A., Boguski, M., Lipman, D.J., and Ostell, J. (1994). GenBank. *Nucleic Acid Res.* 22, 3441-3444.
 McGarvey, P.B., Huang, H., Barker, W.C., Orcutt, B.C.,

- Garavelli, J.S., Srinivasarao, G.Y., Yeh, L.L., Xiao, C., and Wu, C.H. (2000). PIR: a new resource for bioinformatics. *Bioinformatics* 16, 290-291.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235-242.
- Jeon, Y.J., Oh, J.H., Yang, J.O., and Kim, N.S. (2005). KBUD: The Korea Brain UniGene Database. *Genomics & Informatics* 3, 86-93.
- Kim, Y.U., Kim, I.H., Bang, O.S., and Kim, Y.J. (2008). StrokeBase: a database of cerebrovascular disease-related candidate genes. *Genomics & Informatics* 6, 153-156.