

Accessing and distributing EMBL data using CORBA (common object request broker architecture)

ABSTRACT

By integrating CORBA into existing systems and providing users with built-in clients for their applications, we can provide a more efficient way to access EMBL data while reducing the challenges of traditional flat-file formats.

INTRODUCTION

In the field of medicine, we have a number of different ways of accessing medical data. EMBL is an open-source data set that was developed by the EMBL Institute and has been used by thousands of medical researchers worldwide. The use of EMBL data is open and transparent, and the system is designed to be easy to use and maintain.

By understanding the structure of the data architecture, you can help the system manage data management, and increase the efficiency of the data handling by enabling the use of the CORBA (common object request broker architecture) data communication protocol.

The data is stored in a table called the 'Table of Contents', and the data is sent over the network using the 'GET' method which is a data transfer protocol.

In this article, we will create The EMBL database, which is commonly known as the EBJ database and is located at EBI, is a vast collection of DNA and RNA sequences that are submitted directly by researchers and genome sequencing groups. It is obtained from scientific literature and patent applications. Each of these groups collaborates with GenBank and DDBJ to collect 80% of the global sequence data, and all new and updated database entries are exchanged daily. As scientific knowledge grows, the complexity of related information increases and the structure of the data changes. The EMBL database is managed by Oracle, a relational database management system (DBMS), which has over 130 tables and 140 relationships, with approximately 10 million primary data objects and sub-objects. It is available in flat-file format through ftp, CD-ROM, tools, and other means, but there are limitations to using flat files, including difficulties in representing and handling complex data, difficulty in extracting meaningful units, struggle to assembling objects, etc. OMG's CORBA architecture is a standard architectural framework for portable distributed object-oriented computing applications. It provides platform-independent programming interfaces and models that are free from programming languages, computing platforms, and network protocols, making it an ideal solution for creating new applications that can query and distribute biological data.

CONCLUSION

Remarkable conclusions This article discusses the development of a CORBA infrastructure at EMBL-EBI, which uses the resulting MLFM (Object Model for Logical Operators) to create the CORDBA server. PersistenceTM is used to map the object model to the relational schema of the Oracle database, while properly constructed loaders use the "live object caching" technique to increase performance. The evictor pattern is utilized for memory management. The next step in the ongoing work will be

to move the implementation of the EMBL server to meet the new OMG standard for biosequences, as per the OMR rules. By applying these rules, the EBI, being a co-submitter on the Biomolecular Sequence Analysis (BSA) standard, is required to implement the standard. As the BSA proposal is not fully compatible with the IDL specification currently used, care will need to be taken to make this transition as easy as possible for existing clients. Further details are available. Supplementary information, such as The EMBL Nucleotide Sequence Database object model and The International Labour Organization (ILO) specification, is included in the online version of this article.