A Workbench for the Research and Development of Biological Diversity

Hideaki Sugawara hsugawar@genes.nig.ac.jp
Satoru Miyazaki smiyazak@genes.nig.ac.jp

Center for Information Biology, National Insatiate of Genetics,
1111 Yata, Mishima, Shizuoka 411-8540 , Japan

Keywords: biological diversity, workbench, CORBA, XML

1 Introduction

Experts believe that we have recognized only about 10% of nearly 14,000,000 living species. This scale of biological diversity (biodiversity) will be one of major targets of the research and development in the 21st century. The Convention on Biological Diversity (CBD) is the driver of this tendency. The genome projects are another driver because they give us a full set of genes, which make a species.

In the age of biodiversity, classification and identification are key procedures to organize new species into taxonomy. We developed and report here a workbench that promotes the research and development of biodiversity from the standpoint view of the systematics.

2 Method and Results

We developed the workbench by use of data of yeast cultures that were kindly provided by Dr. Takashi Nakase of Japan Collection of Microorganisms (JCM) in the Insatiate of Physical and Chemical Research (RIKEN), Japan.

The hardware for the workbench is DELL Dimension PC with 450 Mhz×1CPU model and the operating system is Linux. We applied the Common Object Request Broker Architecture (CORBA) to the workbench for the interoperability with other databases on biological resources. We selected ObjectSotre, JDBC and ORBacus to construct the database and to implement CORBA. We write the application programs for the analysis and interactive user interfaces in JAVA language to realize Write Once, Run Anywhere. We were also careful to make the workbench modular.

We could develop the workbench that has three features in the following:

1. The workbench is machine-independent.

2. It captures data from distributed databases on the Internet to store them in the local database. Fig. 1a is an example of storing data from 4 different databases that are shown in Fig. 2.

3. It is easy to modify the database schema and to cope with addition of new fields for tests and updating scientific names

4. It provides many views for the comparison of results given by different analytical methods and different category data from the same set of organisms. Fig. 3 is a sample session of analyzing data by use of cluster analysis based on phnotypes and phyogenetics analysis based on sequence data.

5. The workbench suggests the user of the additional information to improve the classification and identification.

We evaluated XML technology while we developed the current version of the workbench. Based on the evaluation, we will convert the database engine to XML files in order to improve the portability of the system especially for the biological resources centers that often lack powerful information environment and informartics experts.

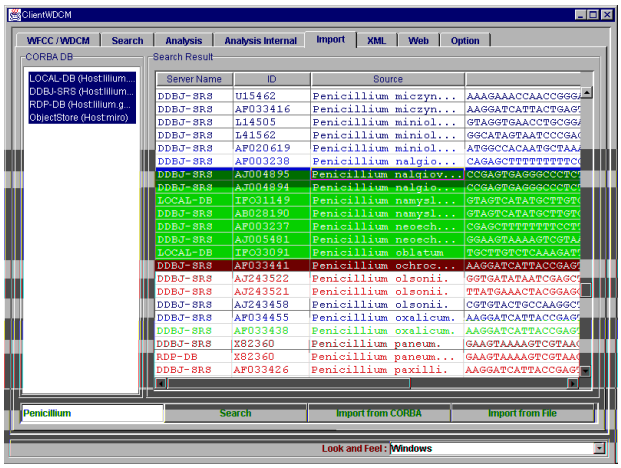

Figure 1: An example of storing data from 4 different databases

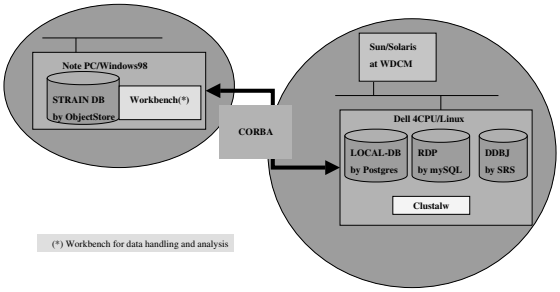

Figure 2: 4 different databases

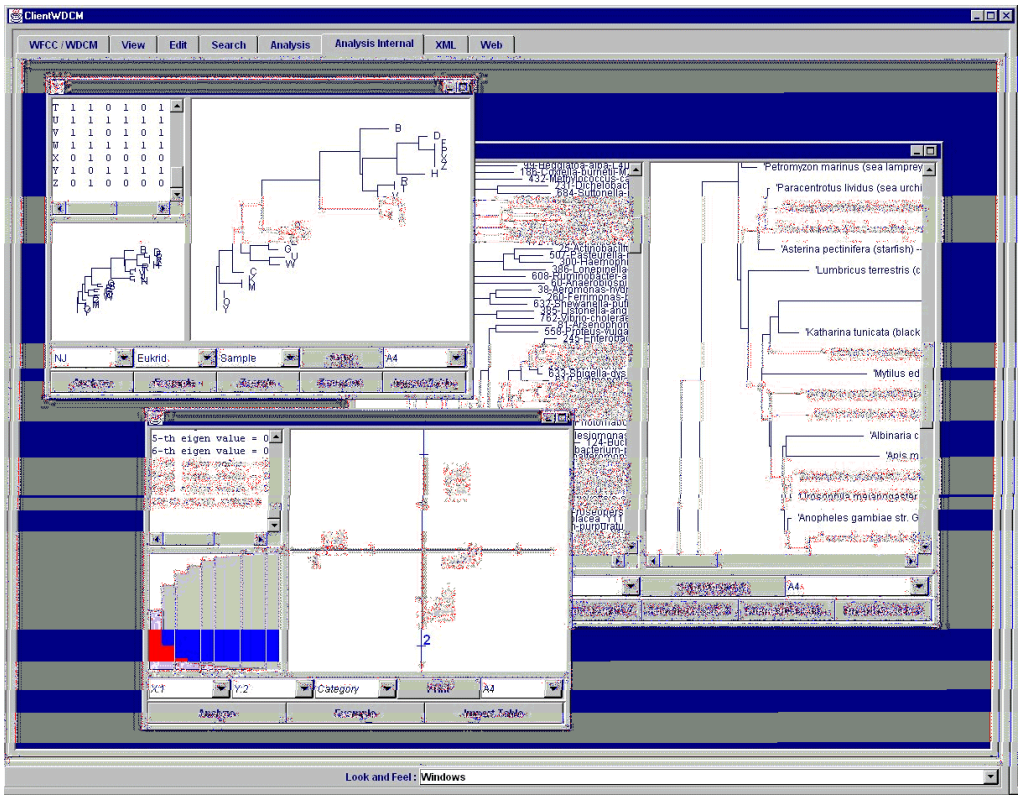

Figure 3: A sample session of analyzing data by use of cluster analysis