Development of a genetic information system for institute engaged in molecular biology.

Hiroshi Mizushima  Yosihyuki Kuchino
hmizushi@gan.ncc.go.jp ykuchino@gan.ncc.go.jp

Biophysics Division
National Cancer Center Research Institute.
5-1-1 Tsukiji, Chuo-ku, Tokyo 104, Japan

Abstract

National Cancer Center Research Institute has about 200 researchers. Most of them are engaged in studies of molecular biology. Before the introduction of Local Area Network system (LAN), each division in our institute had their own genetic analysis system working independently. As we integrated LAN and connected it to the internet by joining TISN, we planned to make a total system for genetic analysis. Our concept was to make a system so that all researchers can use updated versions of database and analyze their data by supercomputers, easily. We thought it is hard for all researchers to use UNIX commands to perform their searches and retrievals to the database. So, our system is consisted with PC-9801 or Macintosh based system. It uses daily-updated database from NCBI to search by PC-9801 or macintosh programs. It includes gopher system to retrieve required sequence from database, and connection to mail-servers at many places to perform homology searches using supercomputers. As we are maintaining a gopher server in our institute, we are now making a gopher interface to many databases to retrieve and search data easily. We are performing many services not only to the researchers in our institute but also to the researchers outside our institute.
1 Introduction

The information about genome or DNA sequence has been growing so rapidly that biologists are required to access to the huge databases and also analyze their data with big computers. However, they are not familiar to big computers and complicated systems. Also, most of the biologists are not familiar to main-frame computers, host computers or even workstations.

Computer systems is changing from host type to server-client type using network systems. High performance workstations are available in low price these days. Also, computer network systems are expanding widely, and many services are available on internet.

As we introduced our local genetic information system in this meeting last year, we established a LAN system in our institute. We also made internet connection by taking part in TISN project. Using this environment, we established system as described in this report. We also started many services to outside our institute.

2 Methods and Results

2.1 Local Area Network System

In national cancer center, we now have more than 300 instruments connected to our LAN as shown in Table.1.

<table>
<thead>
<tr>
<th>Table.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macintosh computers</td>
</tr>
<tr>
<td>Macintosh printers</td>
</tr>
<tr>
<td>NEC PC-9801</td>
</tr>
<tr>
<td>IBM PC</td>
</tr>
<tr>
<td>UNIX WS</td>
</tr>
<tr>
<td>HUB etc.</td>
</tr>
</tbody>
</table>

Most of the researchers are using personal computers and CD-ROMs for their analysis in each division. There were only few researchers who were using computer centers for their genetic analysis. I also thought it was hard to teach all researchers how to use UNIX commands and some new programs. For that reason, I decided to use a Macintosh or PC-9801 computer based system on the network.
2.2 Local genetic analysis system.

Total scheme is shown in Figure 1. We chose a commercial software (GENETYX) which provides a genetic analysis system for PC-9801, macintosh and UNIX workstations. The database for theses system had same structure. We placed this software on Netware file server so that PC-9801 users and macintosh users can use same database systems. Each researchers don't have to care about the database update because it is done on the server, not on each local CD-ROMs. We also have a daily update of the NCBI database on our local UNIX server through the internet, and developed a news system so that these daily-updated data are converted to the database format on the Netware file server.

Some researchers are not sufficient with this system about the speed for the homology search, because it is done by the CPU in local personal computers. For this problem, we are using Public Domain Software(PDS) called Eudora on Macintosh computers so that each researchers can send e-mail to homology search mail servers without using UNIX mailing commands or filing systems. We are planning to make a system to send a e-mail to the mail servers directly from the genetic analysis system itself automatically.

To get some other information, such as GDB, PDB, TFD, etc., we are using gopher system. This system is very useful and fast to retrieve and search for some information from its host machine. It also gets image data from the database such as 3D structure of a protein.

We also registered as the NCBI entrez network user, so that each macintosh or UNIX user can now use NCBI network entrez to search and retrieve the data. We are using a PDS called GDB-accessor to access GDB. This software is very easy to retrieve the data.

2.3 Using gopher system for genetic research using internet.

We started the gopher server at our institute and implementing many information on it. We have cancer related database(CancerNet), cell-bank and gene-bank catalog, information of Japan, and other databases on our gopher server. In our server, I also gathered search and retrieval menu for DDBJ, genbank, PIR, TFD, EPD, GDB,OMIM databases. These searches are a link to each database at their host computer. I am now making a daily updated database on our sever to answer these searches. The usage of our server is about 2000-3000 queries per day. Judging from our log file about the usage of our server, most of the users are from U.S.A.(60%). There are only 20% users from Japan and 20% are from other countries. These data suggests that the internet is well distributed in U.S.A. and not enough established in Japan. There are many people who wants to use gopher networks or mail system, however they cannot join internet society.

2.4 Service for users outside National Cancer Center.

Our gopher server is opened to the world, and can be used by the gopher client software. For
those who does not have gopher client software, we have a telnettable gopher server, so that anyone on the internet can login to our system and use it. It is available by telnetting to gopher.ncc.go.jp and login as gopher from a vt100 terminal. For those who is not on the internet, but can use e-mail system, gopher mail server is available. It is not comfortable as online use, but same information can be taken. Mail to gophermail@ncc.go.jp with help in the Subject:

Another service is a news and mailing list conjugated system. There are bionet.* newsgroups and fj.sci.bio newsgroup to discuss about biological topics. There are Comp-Bio mailing list discussing about computer related biology. There are also many discussion at BioNet forum at Nifty-serve (PC-NET). These system are individually discussing and there are no crosstalk between them. We
thought there should be a global discussion group which anyone can attend. To overcome this point, we started a trial system which mails the message coming by news system. Also we made a mail-to-news system so that anyone can post news by e-mail. Using this system, those who can use e-mail or news system can discuss in a same system including PC-network such as Nifty serve.

3 Discussion

Theses kind of information system is very useful for general molecular biologist, but it may be not enough for laboratories engaged in human genome project. However, for these laboratories, many specialized software are established in genome-informatics project in Japan.

Also, this requires internet connection. Institutes attending on genome-informatics project can easily join GeneomeNet, however, it is very hard to join internet for people not accepted to join genome-informatics project, or university computer center network systems. This is a big problem in internet project. There should be a system so that any academic institute can join the internet society easily without any limitations, although there should be a budget to maintain the internet backbone.