embnet.news

Volume 1 Nr.2 (ISSN 1023-4144) December 23, 1994

Welcome to the second edition of embnet.news! This newsletter is intended for all researchers who are interested in, or have a need to work with molecular biology data on computers. This newsletter naturally focuses on, but does not limit its scope to reports and new achievements from the European Molecular Biology network (EMBnet).

We received very positive feedback on the first edition of this newsletter and would like to thank all readers for their interest. Biocomputing matters face an interesting challenge these days: Positioned between the actual experiment and the writing of conclusions, the researcher has to use methods which are not neccesarily well characterised and known to him. embnet.news provides a forum of communication which will benefit the users, providers and developers of biocomputing methods.

The need to learn

There is a vast majority of methods known in molecular biology lab work. Similarly, a lot of methods are available to exploit the results of the lab work on the computer. However, in contrast to the 'wet' techniques, biocomputing methods are not used every day and therefore the selection of methods becomes rather stochastic. 'Popular' tools are reused quite often, even if scope and limitation of a related program would be better suited to answer the particular query.

New methods are frequently applied rather than investigated before use. The basics of networking, for example, are acquired on an 'ad-hoc' basis by starting a program which performs a task, rather than getting an introduction in the lecture hall and proceeding more effectively thereafter. The BITS article - BioInformatics Theory Section - focuses on the basics of client/server computing.

The need to develop new tools

Biological data are currently increasing at an exponential rate. The number of users, data sets and methods required to tackle sophisticated questions is increasing dramatically. Present technology, however, will hit limits easily. Take the recent sequence submission of a DNA sequence with a length of about 650000 bp as an example: This, the longest sequence made available so far, has been known to make some non-dynamic programs fail.

Organizing the knowledge

Methods and data become very complex and dedicated languages are required to sort out and help organise the subject and target areas of a particular work. Keywords help tremendously to classify information, and make retrievable information stuck in previously heterogenous descriptions. One of our research reports in this issue focuses on the development of keywords in DNA databases.

Another example which requires new tools concerns the fact that the size of the databases is exceeding the capacity of a single CD-ROM. Whereas previously it was thought that just installing another drive cures the problem, the number of useful databases is increasing as well and often a necessary escape is to utilise client/server technology to access resources via a network. This edition of embnet.news deals with two such developments in the research reports; both asynchronous searching and database browsing are described.

The need to know about it

Many questions on public bulletin boards reflect the need to learn about useful resources. EMBnet is committed to serve as information broker. EMBnet welcomes any molecular biology queries from their users. This is particularly important as, with the vast diversity of software available, people may use software or data not appropriate to the task they wish to perform. This implies not only access to databases but also other services. embnet.news features one of the EMBnet nodes in each issue, and also presents achievements from various nodes in the 'Node News' section.

Other regular features of embnet.news include conference announcements and, as no-one has yet programmed the

Contents

Editorial - Welcome to embnet.news	1
BITS - Client/Server Principle Revealed	2
Node Focus - INSERM, France	3
Database and Program Development	
The EMBL Keyword Hierarchical Structure	4
The Sequence Retrieval System (SRS) on the WWW	5
Biocomputing on a server network	6
Tips from the computer room	8
Node News	9
Conferences	11
The EMBnet nodes	12
Dear Reader	13
Leisure Page	14

perfect interface, there is a corner for tips from the computer room to help biologists get to grips with computers.

EMBnet's role

The European Molecular Biology network recently formed an organization in the form of a Dutch Stichting. An Executive Board and three Project Committees are involved in organization and management of the organisation. Besides administrative issues the Project Committees are currently launching calls for proposals. EMBnet nodes are invited to launch or boost common projects dedicated to teaching, training, service and development activities. This newsletter will keep you updated on these activities. For example, both the SRS and the DNA keyword project have received support from EMBnet funds.

embnet.news ...

is available electronically on 'WWW servers', and as postscript file which you can print locally. It is a platform of information and reports, hopefully both useful and interesting, for researchers using computers in molecular biology. We welcome continuous feedback and appreciate submissions for inclusion in future publications (email address: emb-pub@dl.ac.uk).

Let's make communication a success!

The embnet.news editorial board

Alan Bleasby (SEQNET, UK)
Reinhard Doelz (Biozentrum, Switzerland)
Robert Herzog (BEN, Belgium)
Rodrigo Lopez (BiO, Norway)

BITS
BioInformatics Theory Section

CLIENT/SERVER PRINCIPLE REVEALED

R. Doelz, Swiss EMBnet node

The rapid progress in networking technology benefits various areas in biocomputing. Sequence database searching or retrieval applications, as well as many other useful services, are no longer necessarily stored on a local computer but rather can be used via electronic networks.

The typical work process in the past: Local Work

The most common way to connect to a computer which runs the software you want is to sit at a suitable terminal, Macintosh or PC and enter some kind of address, a username and a secret password. This allows the remote computer to place you into a reserved individual disc area in order to store data to be used or produced by computer programs. For a sequence search, you start a suitable program (e.g. sequence editor) and type in the sequence, or transfer for example, a data file from another computer (e.g., a sequenator). Then you launch the search, and the output of the program will be stored on your disc.

You can do it NOW: Using the network

Ideally a networked program is launched like any other. The use is therefore transparent to you. This implies that there is some hidden mechanism which is capable of finding the address, and authenticating you as a legal user of the remote service. Next, all data needed for execution are transferred to the remote system, the requested program executes and returns the results transparently via the network. This procedure implies that your networked program operates a sophisticated mechanism to do all the networking automatically. The language which is used to talk across networks varies and depends on the needs of the procedure. It is therefore called a protocol.

The Server component

The program which resides at the remote computer will always be ready and waiting for tasks transmitted by the protocol. It is therefore called a server. The remote computer will have the server run under a special user name which allows incoming tasks to be processed as if these were requests typed by someone using their local system. In contrast to the local processing, however, the server will delete all input and output data after having retransmitted the results. As you will usually not need to identify yourself by giving a user name and a password, the server will deal with all incoming requests as independent queries, and therefore allocate a temporary scratch space for each result. The server therefore, will not know if you resubmit the query and does not need to keep results for this reason.

The Client component

Your networked program will need to talk to the remote server with a defined protocol. As different protocols exist for different services offered by different servers you will need to have several programs available which are capable of talking to the remote servers in the right language and launching a request. The programs are therefore called clients . On EMBnet nodes clients are usually installed system-wide, which means that you can use them without having to install them yourself.

Using Clients to access network services

There are two issues with respect to usage of your networked client. As outlined, you need an address and an authentication to be eligible for remote services.

Authentication

Servers usually offer anonymous access, which means that you may use the remote services without requiring an account from the service provider. This is a privilege, and you should not abuse your client if it is possible to do you work locally. The rules of good behaviour on the network are commonly known as netiquette.

Resource discovery

Your Client will be configured to talk to one of the servers which are available on the international networks. Depending on which information you are looking for, or which application you have in mind to be executed by the server, it is important to note that not all servers offer all data or all programs. In order to connect to a specific server you either have to know the address (or name) of the server, or the server will be addressed automatically (as the HASSLE protocol does). Some popular protocols, such as the famous World Wide Web (see below), offer tables of services, so-called meta indexes , which are long lists of server names associated to a given service. Within these lists a simple RETURN or mouse click is sufficient to make your client move to the advertised place.

Using suitable information, you will find services on the network which answer your questions in the desired detail which you could not have imagined before.

Examples

There are several categories of client/server applications. Some are focussed on computer or network maintenance, some are dedicated to issues raised by the end user. If you have used the program telnet before, then you have already taken advantage of an application which requires username/ password authentication. Electronic Mail is also a client/ server system, and does not require that you provide a username and a password on the computer which your colleague uses. Within EMBnet several specific networking protocols are being used in production work.

For example, the Network Data Transfer (NDT) protocol (Peter Gad, EMBnet Sweden) takes care of the data distribution of EMBL data library updates. The Hierarchical Access System for Sequence Libraries in Europe (HASSLE) (Reinhard Doelz, EMBnet Switzerland) lets you do sequence searches and retrieval access the network. EMBnet funds resources to support the development of a RPC application by Klaus Heumann, EMBnet Germany (MIPS), to allow PIR International data to be distributed via the networks. Other popular protocols are being developed elsewhere. The NCBI in the US offers access to the Network Entrez application or the Network BLAST application within a client/server application.
If you read this newsletter using the World Wide Web then you also use a client/server application. The searching of archives, for example the EMBL data library, is possible using an information retrieval system called Wide Area Information Service (WAIS). Another famous information browser is GOPHER which allows to view information as if it were a directory of data files. Most EMBnet nodes run GOPHER or WWW servers and provide access to other such services worldwide.

A very recent and important development is the creation of a network of World Wide Web servers running the SRS system (SRS by T.Etzold, Heidelberg) - Browse more than 40 databases in seconds. Try it out !

NODE FOCUS

EMBnet France
Service de BioInformatique
CNRS-INSERM
7 rue Guy Moquet - BP8
94801 VILLEJUIF Cedex, France

General Information

The primary role of the Service de BioInformatique is to maintain up-to-date molecular sequence databases and computational analysis tools with easy access for users in the research community. This service, which is a joint unit between CNRS and INSERM, has been financially supported by GREG (national genome program). This EMBnet node is also the French GDB node (Genome Data Base).

The present server is located at the SC5-INSERM at Villejuif (south of Paris). Associated with the SC11 INSERM it will constitute an important focus for bioinformatics and genome programmes. Currently we operate on a bi-processor SUN SPARC Centre 2000 and two SPARC 10 workstations; with more than 12 Gb of disc, and communication with RENATER (French IP network).

Contact : Philippe Dessen
dessen@genome.vjf.inserm.fr
Tel : +33 1 45 59 52 41
Fax : +33 1 45 59 52 50

How to apply : Any researcher in France is encouraged to apply for an account.

IP address :192.134.32.20
IP name : genome.vjf.inserm.fr

Access for users: Internet, X25, Modem.

Operating system: SUN Solaris 2.3

Main software:
GCG,ACNUC,SRS,FASTA,BLAST,STADEN, PHYLIP, ...
Main sequence databases : EMBL, GenBank, SwissProt, PIR, OWL, PRODOM, SBASE, dbEST, HOVERGEN ...
Main genome databases : GDB, OMIM, GENDIAG, IGD, ACeDB, AAtDB, AScDB QUICKMAP, EMG ...
Current users : 700+
Funding : national
Charges : free for academic users, only communication costs
Commercial users : We are not currently prepared to deal with those.
Menus: yes (with documentation)
X-windows, EMBnet news, Gopher, Wais, ftp and WWW are supported.

DATABASE AND PROGRAM DEVELOPMENT

THE EMBL KEYWORDS HIERARCHICAL STRUCTURE

M. Attimonelli, F. Licciulli, M. Simone and M. Vitale. Italian EMBnet node

Introduction

The Nucleic Acids Sequence Databases including the EMBL Data library are important in EMBnet activities. Hence any improvement of the quality of the EMBL Data library content benefits the entire EMBnet community. The usefulness of any database is related to its association with an efficient retrieval system. A widely used search criterion is by KEYWORD.

EMBL Data library Keywords are words illustrating the biological function of the sequences they are associated with, but the keywords in the EMBL Data library are sometimes unstructures and consequently, like other databases, contain occasional mistakes and inconsistencies.

Approach

We have tackled the need to organise the keywords into a hierarchical structure allowing their classification based on biological function criteria. In order to represent the hierarchical organisation of the keywords a tree-like structure was chosen. This structure is characterised by a set of elements, called nodes, connected to form a father-son link. One particular element, called root, is placed at the highest level in the tree which is the origin to link to all the branches. To provide an organised structure, new keywords have been introduced as key nodes in the tree.

Detailed description

Selected key nodes are:

PROTEIN
DNA
RNA
UNCLASSIFIED

These appear as direct branches stemming from the tree root and they are the ancestors to all the keywords present in the database.

In the EMBL Data library biologically similar entries are often present with different associated keywords to them. This is due to a lack of standardization in keyword annotation by the researcher (i.e., from the principal authors of the entries). Synonymous keywords are seen both in terms of function and vocabulary. In order to make up for this non-standardised information which renders the retrieval of a complete set of data difficult, we started to produce physical relationships between synonymous keywords.

The link of the hierarchical keyword structure with the entries in the EMBL Data library is assured by the association of each keyword with the EMBL entry names and accession numbers of the nucleotide sequences having the biological function illustrated in the keyword. The link is realised starting from an EMBL Data a library Keyword Index . The keywords in the release under consideration (Release 40-September 1994) cover 35712 items.

Implementation

At present the Hierarchical Structure contains 46152 keywords.

Approximately 50% of these keywords have been classified and corrected. Apart from the keywords in EMBL database other keywords have been added to complete the hierarchical structure (mainly at the top levels of the tree). The maximum depth of the tree is 10 levels. Those keywords in the index file which have not yet had their structure determined were isolated for their later classification in the hierarchical structure.

We have been working on retrieval software which has been designed to extract significant information from the created structure and can be utilised for a more efficient interpretation of the data in the EMBL database. The main commands in the retrieval system are:

Browsing of the structural tree;
Selection of entries based on the keywords;
Display of selected items;
Saving of the selection to external files.

The software allows the browsing of complete branches of

the keyword tree. A detailed interrogation and check of the hierarchical organisation of the keywords in the database is possible. The browsing starts from one keyword and the display of both ascendant and descendant keywords.

Usage

Selection is the main command of the retrieval system. It allows you to obtain a set of entries whose nucleotide sequences perform the biological function defined in the keyword used in the selection and also gives the descending keywords. The selected entries are included in lists that can be displayed and processed by other commands of the retrieval software.

The following commands allow the display of the selected data:

NAMES
INDEX
ENTRY

Additionally, the SAVE command allows for the storage of entry names and accession numbers of the selected entries in a file. This option eases the exploitation of the selected entries for further analysis and processing. In particular, entry names can be saved in a format which is readable with the GCG FETCH command.

The prototype keywords hierarchical structure KEYNET can be used remotely by a telnet connection using the following data: host: arba.ba.cnr.it. username: keywords; password: (user e-mail address). However, we are currently working on a WWW server model in order to implement it as a common purpose WWW browser. In order to improve the structure we are developing a program which is capable of searching the structured keywords of the DE and FT lines in the EMBL Data library. This should improve the number of entries linked to each keyword.

Acknowledgments

This work was partially supported through EMBnet by the CEC BRIDGE program, by MPI (Italy) and by Progetto Finalizzato Biotecnologia e Biostrumentazione (CNR, Italy).

THE SEQUENCE RETRIEVAL SYSTEM (SRS) ON THE WORLD WIDE WEB

T. Etzold, EMBL

Introduction

SRS started in 1989 out of frustration. At that time almost all relevant data was contained in a plethora of different flat file libraries or databanks. This situation is not much different today, despite relational and object oriented databases being increasingly used.

The annotation format of these databanks was more or less consistent but almost all databanks had, and still have, different formats. Databanks started cross-referencing each other but there was no tool to use these references. The initial project dealt with sequence databanks only but was soon extended to support all flat file databanks in biology. To date more than 50 databases are supported by the SRS system.

The System

In biology, as well as in other fields, cross-indexing information is the major source of knowledge and understanding. How to devise a cross-referencing strategy that applies to many different types of information is a major challenge. One way of answering this challenge is to create a description language explicitly. The SRS system has such a feature built-in in the form of the ODD (Object Design and Definition) Language . For the specification of a databanks format and file organization ODD provides a common description interface. With these tools at hand an indexing and retrieval system was developed with the following characteristics:

- Support for several different index types (e.g., strings and numbers).

- The possibility to index subentries (e.g., sequence features) so that they can be retrieved independently.

- Cross-references can be indexed to generate links between databanks.

- Indexing makes links bi-directional and allows them to be combined to create a network of databanks in which it is possible to navigate from one databank to any other.

Only simple queries such as the search by accession number, author name, or by sequence length are supported but are extremely fast and can be combined by logical operators.

To allow queries, the entire set of databases must be indexed. The indexing recognises the native databank format and file organization, parses every entry and extracts words and puts them into one of several indexes. Indexes may be created for each individual data field.

The current release works on all major UNIX platforms, VMS and on 32bit DOS and supports almost 50 different databanks such as SwissProt, EMBL, Genbank, PDB. It has two user interfaces: a command line interface and a SRSWWW server.

The SRSWWW server

The World Wide Web server is currently the main interface to SRS. It has unique features that distinguish it from other servers. Almost all HTML (Hyper Text Manipulation Language) files are created by programs just before being

displayed. This ensures that the pages are always up to date. This makes obsolete the requirement of editing HTML files upon the modification, addition or removal of a databank within SRS. Generation of HTML on-the-fly also simplifies the installation on different sites having different sets of databanks.

The protocol used for World Wide Web is called HTTP (Hyper Text Transaction Protocol), and is a simple Query , followed by a stream of replies . The "Server" will die after having replied to a query and therefore does not remember previous queries. Therefore, the HTTP protocol is called stateless.

The SRSWWW server, however, is stateful and achieves this by maintaining a file with the user context. All queries are saved so that they can be re- inspected or combined in query expressions. This is an essential extension to the normal server, but maintains compatibility with the existing browsers.

SRSWWW supports links between databanks both by hypertext links in the entries or by allowing navigation in the network generated by the indexing.

The SRSWWW server is currently installed on 8 EMBnet nodes and one site in the USA. A status page is automatically generated at least once per day that lists all the servers and all databanks served.

If you use the SRS system and have created parsers for publicly available biological databases please contact etzold@embl-heidelberg.de. I would love to add your parsers to the distribution.

Distribution

Current major version: 4.0
anonymous ftp to ftp.embl-heidelberg.de or a ftp.ebi.ac.uk for UNIX or VMS sources .

Public Service:
URL HTTP://www.embl-heidelberg.de/srs/srsc

Reference:

SRS-an indexing and retrieval tool for flat file data libraries. T. Etzold and P. Argos.
CABIOS 9, 1993, pp.49-57
The use of WWW in Biological Research. R. Dolez and T. Etzold. Computer Networks and ISDN Systems 27 (2), 1994. Electronic publication accesible via http://www.elsevier.nl/

Acknowledgments:

Part of the SRS development program has been financed and carried out by grants provided under the Biomed/EU agreement, and project funds from the EMBnet project (BRIDGE/EU agreement). The initial port to the UNIX operating system was done as part of the EMBnet project in Switzerland, funded by the Schweizerische Nationalfonds.

BIOCOMPUTING ON A SERVER NETWORK

R. Doelz, F. Eggenberger, and C. Wadley.
Biocomputing, Biozentrum der Universitaet Basel

Introduction

Computer networks are currently gaining a growing importance in the area of biological sequence data searching and analysis. The established infrastructure on the Internet allows for world-wide connections to data centres which offer the use of programs and data remotely. We have observed an increasing discrepancy between the computer resources available locally and those accessible via networks. As these are developing in a divergent rather than an integrated fashion, the researcher faces two different worlds which are mutually exclusive. Problems and confusion arise if the remote resource is overloaded or fails. The composition of the query, as well as a suitable format of the input data are also factors which could endanger successful computation. This phenomenon is particularly observed within so-called electronic mail servers.

Hierarchical Access System for Sequence Libraries in Europe

We have developed the Hierarchical Access System for Sequence Libraries in Europe (HASSLE) [1] which enables any computer running the necessary client program to connect to one of the nodes which run a HASSLE server. Using HASSLE software allows molecular biologists to work within their well-known environment as the networking is layered below the user interface part of the program. HASSLE automatically uses the nearest server and redirects computational jobs to other remote sites only in cases of overload or failure. This allows biocomputing services to be offered on multiple, equally qualified hosts.

Resource discovery

HASSLE is a communication protocol and does not implement any biological programs itself. Instead, it is the work-horse which figures out where and how a given job e.g. a sequence search, can be executed on the Internet. The most important aspect of resource discovery is based on the concept of a service. If a service is not registered, the server approached initially may either point to a host where the service is known, or to a central information host which knows about all servers and can give the appropriate information, to where an equivalent service is located. This

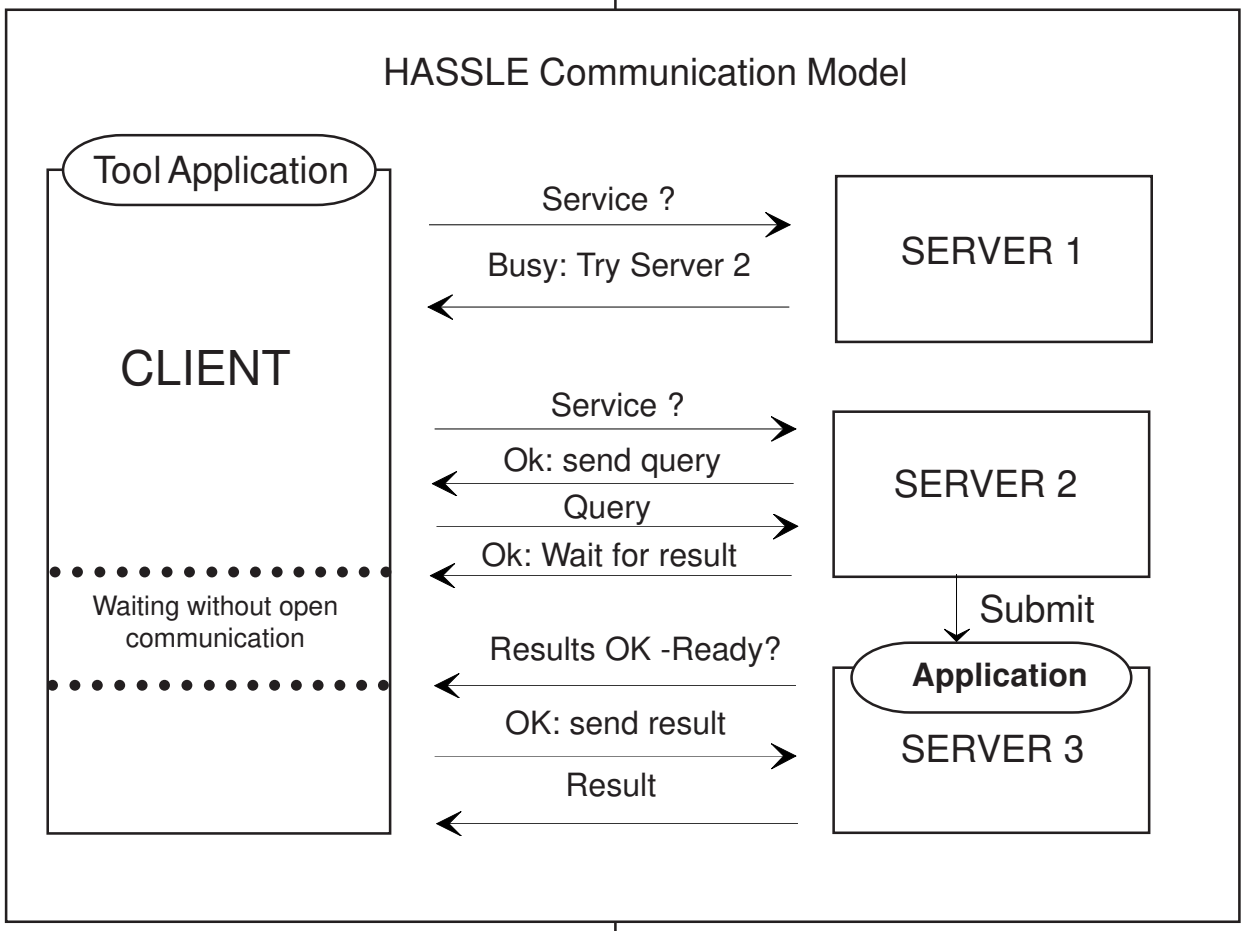


Figure 1: Possible event flow in a HASSLE communication. A Client realises that the configured Server 1 is busy, and delivers the Query to Server 2 automatically. This Server operates in a cluster and possibly resubmits the Query to a batch queue on another host (Server 3), which returns the result after successful completion of the computation.

principal is novel and improves the established concepts of 'links'.

Mode of execution

HASSLE servers may execute the jobs asynchronously in order to allow long computations or a large number of jobs to be executed without the need to have a continuous communication line open. The HASSLE client accounts for this feature by closing the communication after successful job submission and then allowing the server at the remote site to reconnect once the computation is finished. This implementation circumvents a major weakness of the currently available client/server technologies: HASSLE is neither synchronous or bound to a peer-to-peer communication model; the answer to a given question may even come from a different host than the originally approached server (Figure 1).

Authentication

The future of network-based services might require access control in order to account for licensing and general usage limitations. HASSLE implements both access control and security. Regular anonymous access is allowed up to a preconfigurable limit. In addition to the option of the special accounts, all HASSLE communication is encrypted and the use of special encryption keys is possible. Again, access failure causes a redirection as described above in order to achieve a successful job completion automatically.

Usage

HASSLE is now used regularly at many European Molecular Biology Network (EMBnet) nodes. It has proven to be a superior method for access to a network of servers in an integrated environment. The software is freely available. It should be noted that the HASSLE system can be easily modified to be applied in non-biological situations.

Reference

[1] Doelz, R. Hierarchical Access System for Sequence Libraries in Europe (HASSLE): a tool to access sequence databases remotely. CABIOS 10, 31-34 (1994).

Comments and suggestions are most welcome. HASSLE is freely available via anonymous ftp from bioft.unibas.ch in the bioftp-sw directory.

Acknowledgments:

The HASSLE program was supported by Basel University, grants from the Schweizerische Nationalfonds, the Bundesamt fuer Bildung und Wissenschaften and projects within the EMBnet BRIDGE contract.

CompuTips

DOING A MOVE: FILES ARRIVING IN THE UNIX WORLD.

R. Lopez, Norwegian EMBnet node (BiO)

Is your site making the move from good old VMS to UNIX? Are you terrified about what's to happen with all your GCG Fragment Assembly System (FAS) projects? Read this!

First the facts: VMS is not case sensitive. UNIX is.

By tradition, GCG has used VMS's case insensitiveness to decorate their code and the manuals. Words like FindPatterns, StringSearch and PileUp are no longer allowed in the UNIX vocabulary.

The same is true for file names, especially those in the Fragment Assembly Relation directory. If you take a peek at one of these files on a VMS system you will see that the file names begin with a capital letter. You will see Rgh67.seq and not rgh67.seq.

After ftping your project to the new UNIX tin can you will have to fix all the *.fil files in all your relation directories. Up comes the editor? No!!!

The easy fix is in the shell script depicted in listing 1.

Save the script with a nice name like fixrelationfile.sh anywhere on your path and give it execution permissions:

```
% chmod 755 fixrelationfile.sh
```

If you are in favourable terms with your UNIX manager why not ask him to put the script in the /usr/local/bin directory?

Listing 1

```
#!/bin/sh
# fixfasrelfil - fixes uppercase containing filename in GCG FAS relation files
# Rodrigo Lopez 20.05.94.
# to be runned only in the unix relation directory !!!!!!
# print a nice little banner
       echo ""
       echo " FixFasRelFil will replace all occurences of UPPER case letter"
       echo " in a GCG FAS Relation files (filename.fil) that have been moved"
       echo " from a VAX/VMS or OpenVMS system to a UNIX system."
       echo " "
# start a for loop that picks up only the files with the fil extension
        for x in `ls *.fil`; do
# report each filename to the user
               echo "editing " $x
# move each file into another one
               mv $x temp.fil
# use sed's transform command to change all instances of capital to lowercase
# and redirect the result to the original filename.
               sed y/ABCDEFGHIJKLMNOPQRSTUVWXYZ/abcdefghijklmnopqrstuvwxyz/ temp.fil > $x
# remove the temporary file
               rm temp.fil
               done
# done with all files!
# write a little status report to the user.
       echo ""
       echo " Done."
       echo ""
```

Whoops! Before I forget: you must do something about gelenter.init, gelassemble.init and geloverlap.init. These contain the VMS specific slash "/" which should be replaced by a dash "-" in UNIX. Maybe you can modify the above script to fix these files also? Should you decide to delete these files GCG will fetch new ones when you re-start your projects.

Happy gelstarting!

Node News

Belgian EMBnet node (BEN)

Software: SRS installed on both SUN & DEC platforms. GCG 8 and WPI installed

'Current Sequence Awareness' installed; our automatic system to make users aware of update entries.

Databases: Non redundant nucleic acids database from both EMBL and GenBank origins recompiled daily, and accessible through GCG and BLAST

Servers: Gopher server presents WAIS indexed species sections containing integrated sets of "release" and "new" sequences. Gopher also allows searches in complete database in one shot. SRS server via WWW available.

Hardware: Migration to Alpha AXP machine completed.

German EMBnet node (GENIUS)

Summary of recent installed items:

Software:
PHYLIP - phylogenetic analysis programs
STADEN - Roger Staden's package for sequence analysis
Extended GCG
Xpound - predicts coding regions (exons)
Poland - simulates transition curves of RNA/DNA.
Unique - finds 'rare' regions in sequences (e.g. for primer selection)
BLOCKS - protein sequence blocks database
Blimps - compares protein/DNA sequence to the BLOCKS database
MIPSX/PATHX - protein sequence databases and SwissProt updates

ICGEBnet EMBnet node

ICGEB became an Affiliated Centre of the Brookhaven Protein Data Bank.

Biological Computing Division and Israeli National Node

Servers:
A new WWW server has been installed.

Bioccelerator server.

Staff: Dr. J. Prelusky is now in charge of scientific developments and expansions.

Hardware: 15 GB disc space to 30 MB disc space

CNRS-INSERM, France

During the last months, the Bioinformatic node (CNRS-INSERM, Villejuif, France) has installed some new software packages and databases (under Solaris 2.3). It has now more than 700 accounts.

Software:Package GCG (release 8.0)

Integrated Genome Database (IGD) developed at the DKFZ (in a European collaboration)

Servers: Development of WWW servers:
Information
Sequence retrieval System WWW server

Finnish EMBnet node (CSC)

Staff: The new node manager is Heikki Lehvdslaiho now that Rob Harper is at the European Bioinformatics Institute (EBI).

Hardware: Main molbio computer upgraded to 8 R8000 processor PowerOnyx with 2 Gbyte memory.

Databases: CSC is now a member of the Protein Data Bank Service Association (PDBSA) and as an Affiliated Centre is allowed to distribute PDB database on-line.

Software: GCG 8.0 is now in use. In addition to minor upgrades, new programs available are:

Nclever, a text-based NetEntez client
ClustalW, for multiple sequence alignment
Tbob, a BLAST output browser

Manuals for these and all other MolBio programs at CSC are available on-line.

Servers: The CSC Biobox has undergone a major face lift and has been moved to a new address. Update your links to point to http://www.csc.fi/molbio/.

Norwegian EMBnet node (BiO)

Hardware: Expansion of disc space to 10 GB.

Databases:
ECDC replaces ECD
TRANSFAC replaces TFD
Full PDB mirror
new mimmap.dat (for SRS) available via anonymous ftp from ftp.no.embnet.org

Servers:
SRS 4.0 WWW server officially opened
IGD test installation opened.

Biomedical Centre, Uppsala (Sweden)

Databases:
ECDC
CpGIsle
TRANSFAC
EPD installed.

Servers:
experimental WWW server installed.
SRS 4.0 WWW server installed.

Swiss EMBnet node (Biozentrum)

Software Development:
The remote sequence searching software package 'Hierarchical Access System for Sequence Libraries in Europe' (HASSLE) is being constantly improved by Florian Eggenberger and Chris Wadley. We hope to release a HASSLE V5 as a beta test in early 1995.

The JAM (Just Another Metafile) language has been made publicly available on the bioftp server ftp://bioftp.unibas.ch/survival as beta version. This language allows the production of hypertext, LaTEX and Rich Text-formatted documents from ascii-formatted text and has been used to produce the Biocomputing Survival Guide in its second edition (also in beta test).

Teleteaching
Based on previous success, F. Eggenberger launched the second course "WWW Introduction" for biologists. The course is posted on the newly created newsgroup bionet.software.www and its mirrored WWW archive . 18 postings introduce the biologist to the concept of the World-Wide Web and guide anyone who is interested in obtaining and setting up the software required to access the Web. The current posting level is around 15 and expected to be finished early next year. The Biocomputing Courses 1994 have been launched and are still running at one session per week. Accessible on the World-Wide Web or via anonymous ftp (bioftp.unibas.ch), 11 sessions in molecular biology computing focus on the use of biocomputing application, mostly programs of the GCG program package (GCG Inc., Madison, Wisconsin). Currently, about five task sheets and answer documents have been posted.

Network communication:
A virtual reality conference at the BIOMoo located at the Weizmann Institute Rehovot (Israel), introduced the purpose, scope and limitations of the JAM language. Twelve attendees commented on various aspects of portability, versatility and use. The meeting place was the 'EMBnet Cafe', which was introduced by R.Doelz at the Moo with the help of the node staff there.

The SRS WWW server was presented as a paper at the WWW'94 conference in Geneva. The Server as written by Thure Etzold is now available locally running on a World-Wide Web server at URL http://www.ch.embnet.org/srs/srsc.

UK EMBnet node (SEQNET)

Hardware: MassPar and Bioccelerator computers installed.

Conference announcements

The Third International Conference on Intelligent Systems for Molecular Biology July 16-19, 1995 Cambridge, UK.

Organizing Commitee:

Christopher Rawlings, ICRF, UK Dominic Clark, ICRF, UK Russ Altman, Stanford U, USA Lawrence Hunter, NLM, Bethesda, USA Thomas Lengauer, GMD-SCAI, Germany Shoshana Wodak, U.L. Bruxelles, Belgium

Key Dates:
Papers due: Feb 15, 1995.
Replies to authors: Mar 27, 1995.
Revised papers due: Apr 20, 1995.

The Third International Conference on Intelligent Systems for Molecular Biology (ISMB-95) will take place at Robinson College, Cambridge, England during July 16-19th, 1995.

The ISMB conference is intended to bring together scientists who are addressing problems in molecular biology using advanced computational methods including data modelling, machine learning, artificial intelligence, cognitive science, robotics, combinatorial/stochastic optimization, adaptive computing, string & graph algorithms, linguistic methods and parallel computer technologies. The scope extends to any computational method or system supporting a biological task that is algorithmically, cognitively or conceptually challenging, involves a synthesis of heterogeneous information, or in some other way exhibits the emergent properties of an "intelligent system." In general these methods will have been validated on real data sets or have clear practical applications.

ISMB'93 in Bethesda, MD and ISMB'94 in Stanford attracted large and enthusiastic audiences of scientists involved in application areas including molecular biology data and knowledge bases, genetic mapping, gene identification, DNA, RNA and protein sequence analysis and structure prediction, and modelling of biochemical processes. We are continuing the tradition of soliciting original papers which will be rigorously refereed and published (by AAAI Press) in proceedings available at the conference.

The four-day conference will feature introductory and advanced tutorials (on July 16th), and presentations of original refereed papers, posters and invited talks (on July 17th-19th).

Paper submissions will be accepted for review with the understanding that the same work has not been published elsewhere. Papers should be single-spaced, 12 point type, 12 pages maximum including title, abstract, figures, tables, and bibliography with titles. The first page should include the full postal address, electronic mailing address, telephone and FAX number of each author. Also, please list five to ten keywords describing the methods, concepts and problems discussed in the paper. State whether you wish the paper to be considered for oral presentation only, poster presentation only or for either presentation format. Six copies of each paper should be submitted to the address below. For more information see the ISMB'95 World Wide Web page.

WWW URL:
ftp://ftp.icnet.uk/icrf-public/ismb/ismb95.html
email : ismb95@biu.icnet.uk
Fax : +44 (0) 171 269 3067

Please submit papers to:

ISMB'95,
c/o Dr. Christopher Rawlings,
Biomedical Informatics Unit,
61, Lincoln's Inn Fields,
London, WC2A 3PX.
UK.

Travel Fellowships:
Funding has been obtained from the Human Capital and Mobility Programme of the CEC to support fellowships for young (i.e. in post-doctoral and post-graduate research) scientists from EU member states as well as nationals from Austria, Finland, Iceland, Norway and Sweden who wish to attend the conference. Fellowship applications are also open to citisens from other countries, providing they have been conducting research in an EU member state for 12 months or more.

The organisers will be making applications from other national and international funding agencies in the coming months and we hope to be able to announce further fellowship opportunities in due course.

From Toy To Tool - A Symposium on Computer Aided Chemistry and Bioinformatics on the occasion of the 10th anniversary of the CAOS/CAMM Center October 23-24, 1995

You are invited to attend a two day symposium devoted to the developments in Computer Aided Chemistry, on the occasion of the tenth anniversary of the CAOS/CAMM Centre. The full announcement of this conference is available on the Web .

The scientific programme includes:

- Plenary lectures on general computer-aided chemistry topics, in which international experts express their view on the developments in their fields, including data analysis, networking and visualization.

- Parallel research presentations on the main areas covered by the Centre:

Computer Aided Organic Synthesis
Computer Aided Molecular Modeling
Computer Aided MacroMolecular Sequence Analysis/Bioinformatics by international and national speakers.

- A poster session on the above mentioned subjects.
- An interactive display of currently available (chemical) information services in the workplace.

- A Reception and Conference dinner are scheduled for Monday evening.

Date and location:
The symposium will be held on Monday 23 and Tuesday 24 October 1995, at the Faculty of Science, University of Nijmegen, The Netherlands.

Accommodation:
The organisers can make reservations for hotel and hostel accommodation. The second circular will contain the options.

Posters:
A poster session will be part of the program. For computer and/or video presentations you are invited to contact the organisers.

Organizing commitee:
The organisers can be contacted at the CAOS/CAMM Centre,

Mrs. M. van Breemen
CAOS/CAMM Centre
University of Nijmegen
Toernooiveld
6525 ED NIJMEGEN
The Netherlands
Email: secr@caos.kun.nl"

The EMBnet nodes

AT EMBNet (grabner@embdec.bcc.univie.ac.at)
Bio Computing Centre
University of Vienna, Vienna, Austria

BE BEN (rherzog@ulb.ac.be)
Brussels Free Universities,
Rhode-St-Genese, Belgium

CH Biocomputing Basel (info@ch.embnet.org)
Biozentrum der Universitaet,
Basel, Switzerland

CH ROCHE (doran@embl-heidelberg.de)
Hoffmann-La Roche,
Basel, Switzerland

CH SWISSPROT (bairoch@cmu.unige.ch)
Med. Biochem. Dept. CMU, University of Geneva
Geneva, Switzerland

DE EMBL
European Molecular Biology Laboratory,
Heidelberg, Germany

DE GENIUS
(dok419@genius.embnet.dkfz-heidelberg.de)
DKFZ,
Heidelberg, Germany

DE MIPS (mewes@ehpmic.mips.biochem.mpg.de)
Max-Plank-Institut fur Biochemie,
Martinsried, Germany

DK BIOBASE (hum@biobase.aau.dk)
BioBase,
Aarhus, Denmark

ES CNB (carazo@samba.cnb.uam.es)
Centro national de Biotechnologia
CSIC, Madrid, Spain

FI CSC (Heikki.Lehvaslaiho@csc.fi)
Centre for Scientific Computing,
Espoo, Finland

FR CEPH (claude@genethon.fr)
GENETHON,
Evry, France

FR CNRS-INSERM (dessen@genome.vjf.inserm.fr)
INSERM,
Villejuif, France

GB HGMP (mbishop@crc.ac.uk)
Human Genome Mapping Project Resource Centre,
Hinxton, Cambridge, United Kingdom

GR EMBnet Node (savakis@myia.imbb.forth.gr)
Institute of Molecular Biology and Biotechnology,
Heraklion, Greece

HU EMBnet (remenyi@abc.hu)
Agricultural Biotechnology Centre,
Godollo,Hungary

IL INN (isestern@weizmann.weizmann.ac.il)
Weizmann Institute of Science,
Rehovoth, Israel

IT CNR (attimonelli@mvx36.csata.it)
Consiglio Nationale delle Ricerche,
Bari, Italy

IT ICGEB (pongor@genes.icgeb.trieste.it)
International Centre for Genetic Engineering,
Trieste, Italy

IE INCBI (atlloyd@acer.gen.tcd.ie)
Irish National Centre for BioInformatics,
Dublin, Ireland

NL CAOS (jackl@caos.caos.kun.nl)
Katholieke Universiteit,
Nijmegen, Netherlands

NO BiO (rodrigol@biotek.uio.no)
Biotechnology Centre of Oslo,
Oslo, Norway

PL IBB (piotr@ibbrain.ibb.waw.pl)
Institute of Biochemistry and Biophysics,
Polish Academy of Sciences, Warsaw, Poland

PR EMBnet (Pedro Fernandes)
Instituto Gulbenkian de Ciencia,
Oeiras, Portugal

SE EMBnet.se (gad@perrier.embnet.se)
Computing Department, Biomedical Centre,
Uppsala, Sweden

UK EBI (datalib@ebi.ac.uk)
European Bioinformatics Institute
Cambridge, United Kingdom

UK SEQNET (bleasby@daresbury.ac.uk)
EPSRC Daresbury Laboratory,
Daresbury, United Kingdom

Dear Reader,

If you have any comments or suggestions regarding this newsletter we would be very glad to hear from you. If you have a tip you feel we can print in the Tips from the computer room section, please let us know. Submissions for the BITS section are most welcome, but please remember that we cannot extend space beyond two pages per article. Please send your contributions to one of the editors. You may also submit material by Internet E-mail to:

emb-pub@dl.ac.uk

If you had difficulty getting hold of this newsletter, please let us know. We would be only too happy to add your name to our mailing list. This newsletter is also available on-line using any WWW client via the following URLs:

http://www.ch.embnet.org/
http://www.no.embnet.org/
http://www.be.embnet.org/
http://www.uk.embnet.org/

The On-line version (ISSN 1023-4152):

http://www.ch.embnet.org/

A Postscript version (ISSN 1023-4144)is also available, or you can get it by anonymous ftp from:

s-ind2.dl.ac.uk
in the directory pub/embnet.news/

or dbmdec5.ulb.ac.be
in the directory pub/embnet.news/

or ftp.no.embnet.org
in the directory pub/embnet.news/

Publisher:

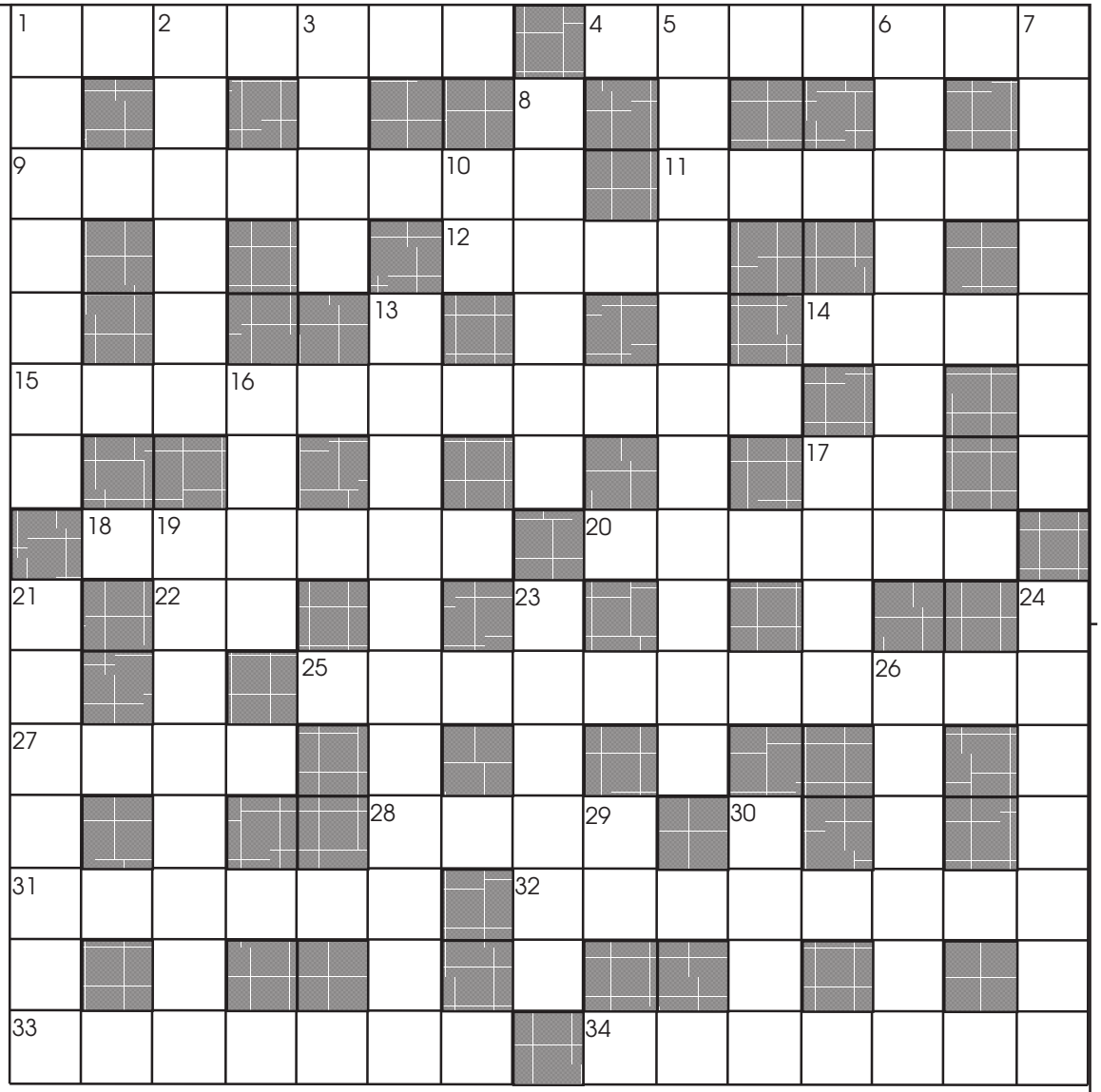
EMBnet Administration Office.
c/o J.Franklin,
ASFRA BV,
Voorhaven 33,
1135 BLEDAM.
The Netherlands

Editorial Board:

Alan Bleasby, SERC Daresbury, UK
(bleasby@daresbury.ac.uk)
FAX +44-925-603100
Tel +44-925-603351
Reinhard Doelz, Basel university, CH
(doelz@comp.bioz.unibas.ch)
FAX +41-61-2672078
Tel +41-61-2672247
Andrew T. LLoyd
Irish National Centre for Bioinformatics
(atlloyd@acer.gen.tcd.ie)
Robert Herzog, Free University Bruxelles, BE
(rherzog@ulb.ac.be)
FAX +32-2-6509767
Tel +32-2-6509762
Rodrigo Lopez, BiO, University of Oslo, NO
(rodrigol@biotek.uio.no)
FAX +47-2-2694130
Tel +47-2-2958756

Across:
1. Most excellent recipe for chicken (7)
4.Frenchman who became an institute (7)
9. Capable of enduring insults(8)
11. Le Boss (6)
12. Where Italy connects (4)
14. Semi-precious UGA (4)
15. Old european (11)
17. Sun god (2)
18. UUU - GCU = (6)
20. Status of most third positions (6)
22. Element associated with Alzheimer's (2)
25. Integer with only two factors (5,6)
27. Small Southern whirlpool (4)
28. Implement (4)
31. Invertebrate found in restaurants (6)
32. Scandinavian Armageddon (8)
33. Snow transport sounds murderous (7)
34. Your status when you've finished this puzzle (2,5)

Down.
1. Recommended position for programmers (7)
2, 3 Most recent additions to EMBnet as they might say (6) (4)
5. Ambitions as natural as breathing (11)
6. What we all are (8)
7. Irritates (7)
8. ATGs (6)
10. Important though short (2)
13. Extremely English..or Welsh..or Scottish (4,7)
16. Le vingt-cinq decembre (4)
17. A lot of paper (4)
19. Partner of software (8)
21. Edam, Gloucester, Blarney, Jarlsberg, Brie (7)
23. Bring in (6)
24. Goes with a bang at christmas (7)
26. Often seen on hippies, goats and mussels (6)
29. The French female musical note (2)
30. Still seen at christmas, despite global warming (4)



embnet.news
ISSN 1023-4144

Given the following sequence, try to translate it into English and take the contents literally.

ACGAGCTTCGTACCATTCGTTGTACGGCGGAGCGTGAGCGTCGTTAGCGG
AAGCCATGGTGGAGATACGGTGGCAGTAACGACGTTCCATGGATTCGTGG
GAGATCCAGGTTTCGTTRTYCATTTC