NETWORKED INFORMATION SERVICES: BRINGING ORDER TO CHAOS?

J FOSTER

Rapporteur: Christyn Emmet

NETWORKED INFORMATION SERVICES: BRINGING ORDER TO THE CHAOS?

Jill Foster
Computing Service
University of Newcastle upon Tyne
NE1 7RU

The networked information environment: the current situation

The Nineties have seen an exponential increase in the number of machines connected to the global internet. Each researcher with his networked PC, Macintosh or Workstation has the potential for "publishing" information to the rest of the world via the network. With networking reaching geographical and research areas never reached before, we see a wide range of Networked Information Services sprouting up all over the Net covering a broad spectrum of both subject matter and quality.

Fortunately to date only a small fraction of the millions of Internet users are publishing on the network, but even so, the problems of information discovery, searching, location and retrieval are already severe. The dream is becoming a nightmare. Computing and networking experts are struggling with the concepts of classifying networked information resources and of uniquely identifying them.

Two quotes (both taken from conversations over dinner at the European Network Services Conference in Pisa last Autumn) sum up quite nicely the situation as regards networked information:

> Peter Deutsch (archie):
> "This [Italian] menu is like using archie: I know there is something here I'd like - but I don't know what it's called."
> Lorcan Dempsey (UK Office for Library Networking):
> "What we have is a flea market; what we need is a department store."

The first quote relates to the lack of standardisation as regards the naming of files of information made available on the network. This is a natural result of the lack of publishing controls. Files are often given short cryptic names that mean something to the owner, but don't give enough information to enable others to judge whether the file would be of interest to them.

The second quote refers to the lack of classification and quality control of the information. It is quite easy for a researcher to spend hours roaming the network looking for information of relevance and to find nothing of interest. This is partly due to the fact that information coverage is patchy, in some areas there is not much of value and in others there is much duplication, but it is also due to the lack of organisation and structure. These are, of course, problems that Librarians have been tackling for hundreds if not thousands of years.

The situation reminds one of Disney's Sorcerer's Apprentice: The enthusiastic computer specialists frantically trying to cope with the situation which is rapidly getting out of hand: writing software tools and re-inventing the card catalogue for networked information resources; the arrival of the Sorcerer (the Librarian) on the scene to the embarrassment of his apprentice in the middle of this chaos.

Fortunately the situation is not quite such a nightmare. A great deal of work has been done in this area and librarians have now entered the scene and are working, often in collaboration with computing service staff and computer scientists, on the problems of classifying and cataloguing networked resources.

Networked Information Retrieval Tools

As the network has grown, along with it there has been an increase in the number of

software tools and applications to navigate the network in order to discover, search for and locate some of the many and varied information resources which are part of the network. This has been the computer network specialist's response to the problem. Within the past two years we have seen a wide spread adoption of tools such as the archie servers, the Wide Area Information Servers (WAIS), the Internet Gopher, and World Wide Web (WWW). The following descriptions, based heavily on documentation provided by the tool developers, cover the NIR tools in widespread use. (Note that network people use the terms "information retrieval" and "archive" in a rather simple sense.)

archie

A large amount of information is available via anonymous ftp servers. With several thousand file archives on the network, it would be a daunting task for a user to contact each to find what was available. archie enables network users to search indexes of the files available on these public file archives. It indexes over two million files on some 1200 public servers. Having located files of potential interest using archie, a user can then retrieve the file using Internet file transfer.

Thus archie is a tool which indexes information and enables information to be discovered and located. It was one of the first of the emerging "networked information retrieval" (NIR) tools.

The Internet Gopher

Gopher represented a giant leap forward in handling networked information. It allows a user to browse information resources on the network and, having found something of interest, to retrieve it using the same tool (gopher), rather than having to use a different tool such as ftp (as with archie). Gopher was also one of the first of the "client/server" NIR tools. The user has a client on his computer and uses this to access gopher servers all over the world. This gopherspace of networked information is presented as a series of hierarchical menus. The information in the menus may be on the same or on different computers. The geographic location of the information is largely transparent to the user. He can even place his own "book marks" in order to go straight to a piece of useful information the next time. Many of the Gopher Servers have been set up by volunteers - mainly computing people, although some networks have employed librarians to organise the information.

Gopher allows access to different types of information, including text, pictures, software, sound and video. It can also provide indexed searches of part of the information base. It is a very powerful but simple tool. Its use has spread very rapidly and its popularity is partly due to its friendly simple to use interface and partly due to the ease of setting up a gopher server to a set of information. A complementary service, Veronica, provides a means of searching a very large number of the gopher menu titles world-wide.

WAIS

Like gopher, WAIS may be accessed by either a WAIS client on the user's computer or via a simple terminal session to a WAIS Server. Unlike gopher, WAIS provides a means of searching the text of documents. WAIS provides access to a set of inverted indexes for collections of information. Typically a user will formulate a query such as "find me items about such-and-such", and WAIS will search the collections the user has specified for suitable items and will return a list of these items ranked according to their "relevance". (This will actually be the number of times the string "such-and-such" appears in the text of the item.) The user has the ability to refine the search by marking certain retrieved items as "relevant" and then asking for "more like these". Queries may be rerun automatically at regular intervals to enable a researcher to keep abreast of the latest developments in his area.

Setting up a WAIS Server for a collection of information requires the building of an

inverted index. As this requires extra effort and computer resources, there are only 400 WAIS servers as opposed to several thousand gopher servers. Note that WAIS makes use of an enhanced version of the standard Search and Retrieve protocol, Z39.50.

World Wide Web

Like Gopher, World Wide Web (WWW) provides easy access to a wide range of information on the network via a client/server model. Unlike the other NIR Tools described here, WWW is based on hypertext. It is possible for the information organiser to set up links from words or phrases in a document to another related document or part of a document elsewhere on the network. For example, the references section of a paper could contain links to the actual (electronic) papers referenced. WWW can provide an extremely powerful tool for finding relevant information, but does require a large amount of effort from someone to set up and maintain the links if its full potential is to be realised.

The above covers the NIR tools in widespread use. There is currently some good cooperation between some of the NIR tool developers. There are now gateways between gopher and WAIS and World Wide Web, so that it is possible to access subsets of the information via different methods.

One of the main problems to date as far as these NIR Tools are concerned (apart from their proliferation) is the fact that most of these Tools have been developed by volunteers. Much of the networked community is relying on these tools to access information all over the world from their desktop, and there has been no funding to support these tools.

The American National Science Foundation (NSF) has made available $1.2M of funding for the Clearing House for Networked Information Discovery and Retrieval (CNIDR). Given the current concerns about the reliance of the community on software tools developed by volunteer effort and the pressure being brought to bear on some of these volunteers by their employers to get back to their "real" work, the setting up of a centre for support is very welcome.

The Networked Information Retrieval Working Group

There are many organisations and associations that have recently begun to focus on the proliferating resources and tools for networked information retrieval (NIR). The Networked Information Retrieval Working Group is a cooperative effort between the Internet Engineering Task Force (IETF) and RARE, the Association of European Research Networks. Its aim is to collect and disseminate information about the tools and to discuss and encourage cooperative development of current and future tools. The NIR Group is co-chaired by Jill Foster (chair of the RARE Information Services and User Support Working Group) and George Brett (Director of the Clearing House for Networked Information Discovery and Retrieval).

This group has put together a draft Status Report on the NIR area. The purpose of this report is to increase the awareness of NIR by bringing together in one place information about the various NIR tools, their developers and other interested organisations. The intention is to make this a "living document" and to update it two to three times a year so that it provides a "snapshot" report on activities in this area.

What is covered?

In the current report there is information on the following NIR tools:

Alex	archie	gopher	Hytelnet	NCSA X-Mosaic
Netfind	Prospero	Soft Pages	Veronica	WAIS
WHOIS	World Wide Web		X.500	

The Consumer Report may be obtained via anonymous ftp from mailbase.ac.uk

file name: pub/lists/nir/files/nir.status.report
The report also covers the major groups working in the NIR area. These will be dealt with in turn.

Groups working in the area of Networked Information Retrieval

The *Coalition for Networked Information* (CNI) is based in North America and was set up by EDUCOM, CARL and CAUSE. It has several working groups including ones concerned with architecture, directories and the TopNode Project (Directory of Directories).

The *Internet Engineering Task Force* also has working groups (WGs) on a wide range of aspects of NIR:

Internet Anonymous File Archive WG
> This group looked at trying to solve the problem of the Italian menu. That is, what was needed in order to provide more information about an information resource in a file archive than just the file name. This group has formally closed, but subsequent discussions on common attribute names and defined values for these are currently involving members from the library community.

Integrated Directory Services WG
> The problems of deploying and managing a distributed Directory Service, regardless of the underlying protocol (X.500, WHOIS++ etc.) are covered by this group.

Uniform Resource Identifiers WG
> This group is concerned with uniquely identifying information resources on the network (very much like an ISBN identifies a particular book) and members of the group have been talking to the Library of Congress and OCLC's USMARC WG. The group has identified the need for:
>> URN: Uniform Resource Name
>>> This indicates the equivalence of two instances of a resource
>> URL: Uniform Resource Locator
>>> This gives location and access details
>> URC: Uniform Resource Citation
>>> All the other useful information (author, size, last updated etc)

Integration of Internet Information Resources WG
> This WG was set up to help integrate the various NIR tools and aid interoperability by defining an architecture. This group also acts as a clearing house for putting other NIR tools and protocols through the IETF standards track. Papers on WWW, Gopher and Z39.50 over TCP are being put forward as draft RFCs (Request for Comments).

WHOIS and Networked Information Lookup Services WG
> The WHOIS++ project aims to develop lightweight Directory Services using simple technology.

RARE Information Services and User Support (ISUS) WG.

RARE Information Services and User Support Working Group is an open pan-European group of volunteers whose members are user support staff, librarians, applications developers and information providers. The work of the group includes liaison between national network support staff, documentation, user group support, networked information services and retrieval tools, training materials and integrated user interfaces. RARE ISUS Task Forces in the NIR area include:

TF: Coordination of NIR Services in Europe
> ISUS WG members have agreed to co-ordinate the European gophers informally and to work on subject specific gophers. This work and discussion is progressing in co-operation with the IETF/CNI working groups, the CNI

TopNode Project, Library of Congress and Australian gopher experts.

TF: User Documentation
User pamphlets and short guides on the main NIR services are being produced.

TF: Training, Publicity and Awareness
This is a joint RARE/IETF WG which is building on the work of the Newcastle based NISP/ITTI Network Training Materials Project. The aim is to collect together existing network training materials, identify the gaps in materials available and produce materials where appropriate. Generic materials are being produced, that may (if required) be tailored to specific sites and/or subject groups. The training is to enable them to make effective use of networked information services.

TF: Multi Media Information Services
There is a joint RARE task force of the ISUS WG and the Interactive multi-media WG which is looking at multi-media information services. This group was formed in November '92. It is looking for funding for pilot projects in this area and for a repository of multi-media information for use in such pilots.

TF: UNITE - User Network Interface to Everything
This Task Force is looking at the requirements for an integrated interface to networked information, be it on-line information files, multi-media information, information on mailing lists, etc. The idea is that the user's access to networked information resources should be fully integrated with his word processor, electronic mail utility etc. (Just as some software has integrated word processor, spreadsheet, database and graphics packages into one). Ideally the user would not be aware that he was using gopher, archie or WAIS as the underlying mechanism. NCSA's X-Mosaic has taken a good stride towards this goal - but there is still scope for quite radical new approaches to this. With X-Mosaic, the user still seems to need to know which underlying tool to use.

The Role of the Librarian in Networked Information Retrieval
There has been much written on the "Virtual Library" or the "Library without Walls". Many Librarians feel threatened by this scenario and by the current widespread availability of networked information. It is important that this new challenge is met actively, and that librarians, rather than standing on the sidelines, capitalise on these new opportunities and are proactive in shaping the future of networked information provision and retrieval.

UK Librarians have been prominent both as a group of network users and as a group discussing potential uses of the network. The JANET User Group for Libraries (JUGL) has been very active from the start. Volunteers run the highly successful BUBL, Bulletin Board for Libraries, which is well known internationally. UKOLN (the UK Office for Library Networking) and the newly formed Information Networking Alliance are other national bodies which will help to gather opinions and to formulate policies.

Reference librarians should become involved in training their users to navigate the ocean of networked information. They need to help them develop information retrieval skills to search for information of relevance to them. They should know the networked information in their subject just as well as the printed information - and just as they help users to find printed information they should help and train them to find on-line information.

Much computer stored text is deleted once out of date. For some information, traditional archival services are required. This area needs more investigation. There is a need too for more library and information scientists to become involved in trying to solve the problems associated with organising the world of networked information.

In short, librarians need to take a leading role in shaping the future virtual library.

The new generation of network users

The global Internet has entered a new age. It is now a useful and vital resource for a wide range of groups in addition to computing experts. Mitch Kapor of the Electronic Frontier Foundation has likened the Internet to a wild land inhabited by natives who have adapted well to the environment and learned to cope with the complexities and peculiarities of using the network. He says that the Internet is now being colonised by new groups of users who are more demanding. These new electronic communities will cooperate over the network to solve their problems and exchange information, but they have little time to find out how to use the network. They need professionally run services, support and training and better tools and interfaces. Currently the patchiness of information coverage, the lack of organisation of the information and the perceived complexity of the user interfaces are further stumbling blocks to the new settlers. Use of the network and group communication facilities should be fully integrated into their desk top applications.

It is time for these new communities to move into the age of "agriculture". They need to develop their own discipline specific information services, subject resource guides and virtual collections. They need the software tools to support these activities. If these groups are to be able to make serious use of the network, they will need quality (refereed) information services professionally run and, almost certainly, charged for. In fact they need a whole range of quality in information services, from the newspaper stand and book store to the library and specialised collections to selective dissemination of information via information "brokers".

Facing the challenge

Development of NIR tools is taking place at breakneck speed. Much work is going on to enable the various tools to inter-operate, and to provide better searching facilities. However the range of information available via the popular NIR tools is fairly restrictive at present and in the main is simple ascii text; search facilities are also still fairly basic from the point of view of the information retrieval experts.

Order out of Chaos?

Computing and network experts and some subject specialists have been making a stab at organising networked information. This is an area which could benefit from the application of traditional library and information science skills; however these should not be applied unquestioningly to the new environment. European Biologists setting up a Biology Gopher have pulled together information of relevance to them which would not go under a strictly applied "Biology" classification. With Gopher it is possible to have the same information appear under several different classifications for the convenience and ease of use of the end user, without the need to hold duplicate copies of the information. This may be a trivial example, but is used to illustrate the fact that we are not, and should not be, trying to create a simple electronic replica of a Library with books on the shelves.

Much thought needs to be given to the issues involved in organising networked information. Does it make sense to reinvent the card catalogue here - or is there another more appropriate method? Is there a way to help the user build his own "library", that is, his own view of the information web? Such a personal collection could be passed on to a new research colleague to help reduce the learning curve. We need experts from all the relevant fields to work together and pool their expertise to produce some innovative solutions to these problems.

Quality information

There is scope for building collections of quality information. Collections could be logical collections, where only links to the information are held, rather than the information itself.

In the current environment there is a need for both the current "flea market" approach and the "department store" approach. In some cases we will require the traditional services of the librarian in building and maintaining collections of information thus adding value to them,

for example by sifting it to keep only high quality (possibly refereed) information. I believe, in the future, researchers will become used to paying for "quality" information. In the meantime we require more pilot activities involving publishers to establish the issues involved, such as copyright, charging, accounting and authentication and to develop and deploy appropriate mechanisms for solving some of the problems.

There is a real role here for publishers: professional people who apply the traditional skills of the publishing world to turning selected electronic information into value added products. Software developers need to provide applications to support the publishing process.

Integrated access to electronic information
To cope with the new generation of users, we need to build better integrated tools for group communication and for networked information retrieval. Having "real users" use the network for "real work" will help to provide some feedback on these tools and help guide their development. So far efforts have concentrated on pulling together the existing tools (for example, X-Mosaic). A radically different approach is needed, perhaps linked more with the type of tasks the user would wish to perform in the "real world".

Multi Media Information Services
The next year will see a national fibre research network in the UK. The next decade will see a radical change in the way information is delivered to the desk top. Multi-Media Information Services and full image document delivery services will be possible in the not too distant future. There is a real need for pilot projects in this area, in order to understand the problems that will arise in trying to synchronise delivery of multi-part multi-media information over the network and to develop appropriate search techniques for video clips, images and sound.

Future "Intelligent" Information Retrieval Services
Prototype "Knowbots" exist that allegedly "roam" the network looking for information of interest to their clients. Developments in artificial intelligence could result in an electronic reference librarian, with which the user would engage in a series of questions and answers. This user agent could then regularly perform a search for information of interest to the user on his behalf. Such a scenario was the subject of a BBC programme by Douglas Adams entitled "Hyperland". This may seem rather futuristic and of course, the human reference librarian will always have an important role to play, but it is important that research projects in this area are proposed and funded.

The Nintendo Generation!
We're seeing the Internet move into a new age, that of agriculture, enterprise and endeavour, as new communities start cultivating their own information services and commercial services develop around the network. Networking has become a multi million dollar business almost overnight. Use of the network is no longer confined to the computer scientist and a handful of other researchers. Network users range from Kindergarten pupils to old age pensioners to the President of the United States, from research scientists to commercial users to ordinary citizens. The type of use made of the network is changing and diversifying rapidly. Will the next age be the age of leisure? Perhaps we will see the network used by the "Nintendo" generation to play distributed multi-player multi-media games. Will "video-dating" over the network become popular? Who knows! The only certainty is that in five to ten years time we will see a new generation of network users using the network in quite a different way. The challenge is to continue to meet the needs of the research communities whilst capitalising on the commercial investment in software and services for the expanded market.

[Note this paper will be presented at the European Network Services Conference in October '93.]

Useful References

Ed Krol:
>   The Whole Internet User's Guide and Catalog.
>   O'Reilly and Associates, Inc.

Jill Foster:
>   "User Support and Information Services in Europe: A Status Report"
>   RARE Technical Report 1, March 1993.

Scott Yanoff:
>   Special Internet Connections compiled by Scott Yanoff
>   anonymous ftp csd4.csd.uwm.edu pub/inet.services.txt

John December:
>   Information Sources: the Internet and Computer-Mediated Communication
>   anonymous ftp: Host: ftp.rpi.edu, file: pub/communications/internet-cmc

Jerome H. Saltzer:
>   Technology, Networks, and the Library of the Year 2000
>   (conference paper)

Jill Foster:
>   Status report on NIR activities, July '93
>   anonymous ftp:Host: mailbase.ac.uk, file: pub/lists/nir/files/nir.status.report

Jill Foster:
>   Report on the IETF Meeting in Amsterdam, July '93
>   anonymous ftp: Host: mailbase.ac.uk, file: pub/lists/wg-isus/files/ietf.07.93

Jill Foster:
>   Status report on RARE ISUS WG activities, July '93
>   anonymous ftp:Host: mailbase.ac.uk, file: pub/lists/wg-isus/files/isus-026

Various:
>   RFCs and Internet Drafts
>   anonymous ftp:Host: src.doc.ic.ac.uk
>   anonymous ftp:Host: nic.nordu.net

Author Information

Jill Foster has worked for the last thirteen years with the Development Group of the Computing Service, University of Newcastle, first as a network systems programmer and now as Manager of the UK NISP/Mailbase Service. She also supervises the NISP/ITTI Network Training Materials project. She holds a first degree in Physics, an MSc. in Solid state Physics and an MSc. in Computing.

She chairs the European RARE Information Services and User Support Working Group and co-chairs the IETF Networked Information Retrieval Working Group and the Network Training Materials Working Group.

Email: Jill.Foster@newcastle.ac.uk

DISCUSSION

Rapporteur: Chrisyin Emmett

Professor Lincoln, referring to an earlier comment by Professor van Rijsbergen that the Communications of the ACM were now available free via a gopher service, said that consideration of authorship, attributions and references should be given higher priority than at present. Mrs Foster said that work on network copyright was currently in progress, and that the issue was fundamentally the same as conventional copyright. She agreed with Professor Lincoln that attributions were also important.

Professor van Rijsbergen said that publishers were frightened and were waiting for others to move. Mrs Foster commented that publishers were now generally willing to make plain text available electronically. Professor Coulouris considered that the current chaotic situation should be tolerated for a while, since people were undecided about what they wanted. There was a danger that effort would be wasted in producing unwanted solutions.

Professor Randell remarked that progress was being made, in that there was some standardisation amongst the things being implemented. Mrs Foster agreed, adding that awareness of the problem was now much higher. Professor Coulouris considered that a fundamental model was required.

Professor Capriz stated that electronic services provide trivial information at present. This was because of the enormous cost of collecting and providing quality information such as that contained in the Encyclopaedia of Physics.

Professor Randell pointed out that much useful information could be gleaned from the FAQ files of newsgroups.

Professor van Rijsbergen commented that the present discussion only applied to existing academic staff. The next generation will want "soft" information and very frequent access.