

Elizabeth Cherhal

cherhal@grenet.fr
Centre Inter-Universitaire
de Calcul de Grenoble (CICG)

Françoise Renzetti,

renzetti@imag.fr

Serge Rouveyrol

rouverol@imag.fr
IMAG / CERSI

INTERNET OU LA RECHERCHE INTERCONNECTÉE

LE BUT de cet article est de faire découvrir aux documentalistes et aux bibliothécaires la richesse documentaire du réseau Internet, et d'indiquer les outils de base permettant d'y accéder.

Les uns et les autres sont convaincus depuis longtemps du caractère indispensable de l'utilisation des réseaux. Dans la communauté scientifique, les réseaux de données permettent en effet aux chercheurs d'envoyer programmes et données sur des super-ordinateurs – afin qu'ils y soient traités –, de récupérer les résultats et d'échanger des informations scientifiques avec des collègues.

Interconnexion de réseaux

Malheureusement, la plupart des réseaux sont des entités indépendantes destinées à satisfaire les

besoins d'un seul groupe, puisque les utilisateurs choisissent des systèmes adaptés à leurs problèmes de communication. Une nouvelle technologie a donc émergé, qui a permis l'interconnexion d'un grand nombre de réseaux physiques distincts et les a fait fonctionner comme un tout coordonné. La technologie de l'interconnexion masque les détails des réseaux physiques et permet aux calculateurs de communiquer indépendamment de leurs connexions physiques.

Depuis vingt ans, les agences gouvernementales américaines ont subventionné des travaux de recherche qui ont permis la réalisation d'une interconnexion couvrant tout le territoire américain grâce à la technologie DARPA (*Defence Advanced Research Projects Agency*). Cette technologie comprend un ensemble de normes réseau spécifiant les détails des communications entre calculateurs ainsi qu'un ensemble de conventions pour interconnecter les réseaux et « router » le trafic. Appelés « suite des protocoles d'interconnexion TCP/IP », ces protocoles peu-

vent être utilisés pour communiquer au sein d'un ensemble quelconque de réseaux interconnectés.

En 1979, la DARPA avait constitué un comité informel, la « Commission de contrôle et de configuration d'Internet » (*Internet Control and Configuration Board*, ou ICCB), pour coordonner et orienter la conception de protocoles et l'évolution de l'architecture d'Internet.

Internet apparut vers 1980, lorsque la DARPA commença à faire évoluer les machines reliées à ses réseaux de recherche vers l'utilisation de nouveaux protocoles. Pour encourager les chercheurs à adopter et à utiliser les nouveaux protocoles, la DARPA en réalisa une version peu coûteuse. Comprenant que les communications deviendraient bientôt une partie vitale de la recherche scientifique, la National Science Foundation joua, aux États-Unis, un rôle actif et provoqua l'extension d'Internet. Début 1985, elle lança un programme pour constituer des réseaux d'accès centrés autour de six gros centres de calcul. En 1986, elle étendit ses efforts en matière de réseaux en subvention-

IMAG : Institut d'informatique et de mathématiques appliquées de Grenoble.
CERSI : Centre d'étude et de recherche en sciences de l'information, ENSSIB.

nant un réseau « grande distance », appelé NSFNET, qui relia tous les centres de calcul entre eux. Aujourd'hui Internet comporte des centaines de réseaux individuels répartis dans le monde. L'expansion rapide a posé des problèmes d'échelle inimaginables lors de la conception initiale et motivé les chercheurs pour imaginer de nouvelles techniques de gestion de grandes quantités de ressources distribuées.

ment avec toute personne disposant d'une entrée sur Internet. Les avantages sont nombreux : communiquer rapidement, sans attendre de joindre la personne au téléphone, simplement, et sans déranger quelqu'un dans son travail : le correspondant lit en effet son courrier quand il le veut et y répond à loisir.

Pour les utilisateurs disposant de Macintosh ou de PC connectés au réseau, des interfaces permettent une

La bibliothèque du Congrès, à Washington, offre le « LC Catalog ».

Ftp

Ftp est un logiciel qui permet le transfert de fichiers entre deux machines connectées au réseau Internet. A partir d'un PC, d'un Macintosh, d'une station Unix, on accède à un serveur distant sur lequel on est identifié au moyen d'un « nom d'utilisateur » et d'un « mot de passe », ou à un des nombreux serveurs *ftp* anonymes – c'est-à-dire ouverts à tous – dans le monde.

L'utilisateur dispose des sous-commandes *ftp*² qui lui permettent d'accéder à de véritables bibliothèques – depuis janvier 1993, on peut, par exemple, rechercher sur le « NASA Langley Technical Report ftp Site » les *Technical memorandums*, ainsi que les autres rapports et traductions techniques effectués à la NASA (Langley) au cours de l'année 1992³, choisir le rapport qui intéresse plus particulièrement, et le recopier.

Internet comporte des centaines de réseaux individuels répartis dans le monde

On peut se faire une idée d'Internet en se représentant le prolongement d'un réseau local au monde entier. Les potentialités sont identiques aux potentialités qu'apporterait un réseau local : messagerie, connexion à une machine distante, transfert de fichiers... Extension gigantesque à une quantité énorme de machines dont la progression elle-même est significative : 213 ordinateurs en 1981, 80 000 en 1989, 313 000 en octobre 1990, 376 000 en janvier 1991, 727 000 en janvier 1992, 1 500 000 en janvier 1993 ! 47 pays répartis sur les sept continents sont connectés. La moyenne du trafic croît de 10 % par mois. De 5 à 10 millions de personnes utilisent Internet.

Les services de base

Trois services fondamentaux sont offerts par Internet : les échanges de courrier (ou « mail »), la possibilité de se connecter à une machine distante par « telnet », le transfert de fichiers, « ftp ».

Le mail

Le mail est le service le plus connu d'Internet. Grâce à des machines passerelles, le système de messagerie permet d'atteindre des usagers disposant d'un autre réseau (Bitnet...) et de communiquer simplement et rapide-

livraison du courrier sur le poste de travail, sans qu'ils aient à se connecter à une machine hôte. Ces interfaces¹ sont du domaine public et ne coûtent rien à l'utilisateur.

Telnet

Par *telnet*, un utilisateur peut se connecter à une autre machine connectée au réseau Internet pour établir une session interactive. Des serveurs commercialisés tel Dialog sont accessibles. On peut ainsi atteindre plus de 400 bibliothèques américaines, européennes, australiennes...

Exemple de la commande telnet

imag(7) telnet dra.com

On accède comme invité à Washington, où l'on obtient l'avertissement suivant :

Data Research Associates, Inc.
Copyright 1991. All Rights Reserved.

Records in this database originating with the Library of Congress are copyrighted by the Library of Congress except within the USA.

This database contains all the cataloging records as distributed by the LC Cataloging Distribution Service as part of their « Complete Service ».

This database is NOT the same as the Library of Congress Catalog.

Les conférences électroniques ou « Usenet news »

Usenet est formé par l'ensemble de machines qui échangent des articles regroupés sous un « label » appelé « newsgroup » ou « group ». Les groupes sont classés selon leur centre d'intérêt, comme l'indique leur nom. Afin que l'ensemble des articles publiés demeurent pertinents à l'objectif primitif fixé, certains groupes sont diffusés par l'intermédiaire du « modérateur ». Les *news* permettent très rapidement d'obtenir

1. Appelés Eudora ou Pop.

2. Sous-commandes *ftp* :

– lister le nom du répertoire sur lequel il se trouve (commande « pwd »),

– lister les noms des fichiers du répertoire (commande « ls »),

– changer de répertoire (commande « cd »),

– prendre un fichier sur la machine distante (commande « get »),

– mettre un fichier sur la machine distante (commande « put »).

3. En se connectant à l'adresse suivante : *techreports.larc.nasa.gov* (ou 128.155.358).

4. Véritable éditeur qui sélectionne les articles.

de l'information comme d'informer la communauté.

Sous Unix, la commande « *m* » (*read news*) présente à l'utilisateur les mises à jour des groupes auxquels il est abonné : *fnet.general* concerne les échanges généraux intéressant la communauté française ; *comp.doc.techreports*, les annonces relatives aux rapports de recherche publiés dans le monde dans le domaine de l'informatique ; *imag.congres*, *imag.mac*, *imag.mediatheque* sont des groupes locaux créés à l'IMAG.

Les *news* représentent un outil

Exemple de news sous Unix

```
imag (3) m
***** 36 unread articles in
fnet.general--read now ? [ynq]
***** 40 unread articles in
comp.doc.techreports--read now ?
[ynq]
***** 1 unread article in
imag.congress--read now ? [ynq]
***** 40 unread articles in
imag.mac--read now ? [ynq]
***** 1 unread article in
imag.mediatheque--read now ?
[now]*
```

*Le choix « ynq » présenté entre crochets propose à l'utilisateur de lire le contenu du groupe (yes), d'aller au groupe suivant, de quitter la commande « m ».

d'information documentaire de premier ordre car ils constituent le dialogue public d'une communauté de spécialistes. Certains groupes à parution régulière deviennent de véritables revues électroniques.

En février 1993, lors du colloque *Les professionnels de l'information scientifique et technique au CNRS*, deux *newsgroups* français ont été créés à l'intention des documentalistes et des bibliothécaires. L'un, « *fr.doc.biblio* », est la tribune professionnelle française ; l'autre, « *fr.doc.magazines* », se propose de diffuser le contenu des sommaires des revues françaises de mathématiques appliquées et d'informatique grâce à une saisie répartie. Ce deuxième groupe « modéré » sollicite vivement les collaborations⁵. Le but est d'alimenter une base des publications françaises, équivalente à

la base australienne qui dépouille les publications en langue anglaise⁶, comme d'entraîner les documentalistes à utiliser Internet.

Les nouveaux outils

Le serveur « Archie »

Ce service permet de retrouver la localisation des fichiers contenus sur l'ensemble des sites *ftp* anonymes maintenus dans le monde. *Archie* retrouve le contenu de 800 « sites d'archives » accueillant près d'un million de fichiers. Il a été conçu comme un moyen d'accéder rapidement aux fichiers que l'on cherche sans savoir à l'avance sur quelle machine ils se trouvent.

Wais

Wais est un nouveau logiciel de recherche d'information sur le réseau Internet.

C'est un logiciel « domaine public » de grand avenir. Il est très puissant et fonctionne selon un mode « client-serveur ». Une partie du logiciel s'exécute sur la machine de l'utilisateur, une autre partie sur la machine

serveur. Il est basé sur le protocole Z 39.50.

Le principe général est le suivant : la base documentaire est constituée d'un ensemble de documents qui peuvent être de types différents (*news*, paragraphes...) ; chaque mot d'un document est indexé sur la « machine serveur » ; les questions posées sur la station de l'utilisateur consistent en un ensemble de mots ; le serveur répond par un ensemble de documents classés selon un « score », correspondant au maximum d'occurrences des mots de la question⁷.

Pour retrouver les serveurs existant sur Internet, on utilise le *Directory of servers*, description de toutes les bases *Wais*. En interrogeant ce

5. Pour les lecteurs intéressés, adresser un courrier électronique à l'adresse suivante : *fr-magazines@imag.fr*

6. *Cs-journal-titles.src.*, base répertoriée dans le *Directory* des serveurs *Wais*.

7. Pour permettre l'interrogation d'une base bibliographique sous *Wais*, il faut générer un fichier séquentiel dans lequel chaque notice est séparée par une ligne blanche ; chaque notice correspond à un document *Wais*. L'INRIA Sophia-Antipolis est le centre pionnier de *Wais* en France et offre, depuis plus d'un an, grâce à cet outil, ses catalogues et les rapports de recherche qu'il publie.

Pour accéder à Archie

Pour accéder à Archie de façon interactive, faire telnet sur l'un des serveurs existants :

```
archie.ans.net (New York, USA)
archie.rutgers.edu (New Jersey, USA)
archie.sura.net (Maryland USA)
archie.unl.edu Nebraska, USA)
archie.mcgill.ca (The first Archie server, in Canada)
archie.funet.fi (Finland)
archie.au (Australia)
archie.doc.ic.ac.uk (Great Britain)
```

Au login et au prompt sur le serveur, entrer "archie" pour se connecter. Un message de bienvenue est annoncé, ainsi que de l'information détaillée sur le projet "archie".

L'utilisateur peut entrer une commande au "prompt : archie>", la commande "help" produira des instructions relatives à l'utilisation des commandes "prog", utilisées pour les requêtes. La commande "set" permettra de contrôler différents aspects des opérations du serveur. Il faut taper "quit" pour quit-

ter le serveur "archie".

Nous donnons ainsi un exemple de session :

```
imag(203) telnet archie.funet.fi
Trying 128.214.6.100 ...
Connected to nic.funet.fi.
Escape character is '^]'.
SunOS UNIX (nic)
login: archie
.....
archie> set maxhits 1 (la recherche est limitée à une réponse).
archie> prog libcat (recherche du fichier libcat).
# matches / % database sear 3 / 2 % / 0 %
Host csn.org (128.138.213.21)
Last updated 17:24 30 Dec 1992
Location: /pub/telnet
FILE rw-rw-rw- 109226 Oct 1 1991 libcat-guide (le fichier libcat est disponible sur la machine csn.org (128.138.213.21))
```

« *Directory of servers* », le client récupère en même temps le descripteur de la base pertinente et peut alors interroger.

Il existe actuellement 400 bases. Les sujets présentés sont variés : de grands corpus, la *Bible*, le *Coran* sont indexés en texte intégral ; on y trouve aussi de nombreuses revues, des catalogues de bibliothèques⁸, et même une base multimédia des disques compacts musicaux de la discothèque de l'INRIA⁹, à Sophia Antipolis !

Gopher

Ce système de recherche d'information, né à l'Université du Minnesota, se présente sous forme de menu guidé. Chaque serveur *Gopher* est lié aux autres serveurs *Gopher*, si bien que l'utilisateur peut passer de l'un à l'autre sans percevoir de changement dans son environnement. Il existe des douzaines de serveurs *Gopher* présentant eux-mêmes des douzaines de « sous-menus ».

Pour localiser les ressources existant sur ces serveurs, on se promène, sans le voir, de l'un à l'autre, en suivant des menus. Initialement développé pour être un serveur d'informations au niveau d'un campus, *Gopher*, grâce à sa facilité d'utilisation, s'est rapidement étendu à tout le réseau Internet. On peut, par son intermédiaire, consulter des bases *Wais*, initialiser des sessions *telnet*, tout cela de façon transparente pour l'utilisateur. Des « clients »¹⁰ existent pour pratiquement toutes les machines.

World-Wide-Web (WWW)

WWW est un puissant système hypertexte de recherche d'information, développé au CERN¹¹ à Genève. Il permet de passer facilement d'un

sujet à l'autre. A la différence d'un outil comme « Hypercard », les liens entre les documents sont établis à travers le réseau. Pour avoir une idée des potentialités de WWW, on peut se connecter par *telnet* au site *info.cern.ch*.

Les trois systèmes, *Wais*, *Gopher* et WWW, sont souvent associés dans les gros systèmes d'information. Si *Gopher* est un système susceptible de gérer des bases de données diversifiées, celles nécessaires à un campus, par exemple, *Wais* est l'outil de

World-Wide-Web (WWW) Exemple de connexion par Telnet

```
telnet info.cern.ch ou telnet [4], November's W3 news [5],
128.141.201.74 (Suisse) Frequently Asked Questions[6].
telnet eles2.njit.edu ou telnet What's out there? [7] subjects [8]]
128.235.1.43 (USA, NJ)
telnet vms.huji.ac.il ou telnet What's out there ? [7] Pointers to the
128.139.4.3. (Israël) world's online information,
telnet info.funet.fi ou telnet subjects [8]
128.214.6.10 W3 servers [9], etc.
```

Nous sommes connectés au CERN :

CERN Information Service
(tty4 on nxoc01)

Overview of the Web GENERAL OVERVIEW

There is no « top » to the World-Wide-Web. You can look at it from many points of view. If you have no other bias, here are some places to start :

by subject [1] A classification by subject of interest. Incomplete but easiest to use.

by type [2] Looking by type of service (access protocol, etc.) may allow to find things if you know what you are looking for.

About WWW [3] About the world-Wide Web global information sharing project

Starting somewhere else

To use a different default page, perhaps one representing your field of interest, see « customizing your home page » [4].

What happened to CERN ?

1-6 Up, <RETURN> for more, Quit, or Help : 3

The World-Wide-Web project
WORLD-WIDE-WEB

The World-Wide-Web (W3) is a wide-area hypermedia [1] information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including a summary [2] of the project, Mailing lists [3], Policy

[4], November's W3 news [5], Frequently Asked Questions[6].

What's out there? [7] subjects [8]]

What's out there ? [7] Pointers to the world's online information, subjects [8]

W3 servers [9], etc.

Help [10] on the browser you are using

Software Products [11] What there is and how to get it

Technical [12] Details of protocols, formats, program internals, etc.

Bibliography [13] Paper documentation on W3 and references.

People [4] A list of some people involved in the project.

1-18, Back, <RETURN> for more, Quit, or Help :

6, Up, <RETURN> for more, Quit, or Help : 8

Subject listing INFORMATION BY SUBJECT

See also arrangements by organization [1] or by service type [2]. Mail www-request@info.cern.ch if you know of online information not in these lists;

Aeronautics Mailing list archive index [3]

Astronomy and Astrophysics Abstract Indexes [4] (down)

Bio Sciences [5] separate list

Computing [6] Separate list

Geography CIA World Fact Book [7], India : Miscellaneous information [8], Thai-Yunnan : Davis collection [9]

Law US Copyright law [10]

Libraries [11] Lists of online catalogues etc.

1-36, Back, Up <RETURN> for more, Quit, or Help.

8. En France, celle du Centre international de rencontres mathématiques à Marseille (CIRM), de l'École normale supérieure de Lyon, de l'IMAG et du CIGG...

9. Institut national de recherche en informatique et en automatique.

10. La partie client du logiciel.

11. Organisation européenne de la recherche nucléaire.

la recherche documentaire multi-bases, multiformats, multimédias, tandis que WWW est un système hypertexte permettant de « naviguer » au sein d'un corpus documentaire¹².

L'accès aux sources d'information

Si vous êtes dans un endroit « câblé », si vous avez un PC ou un Macintosh, il faudra ajouter une carte Ethernet dans celui-ci, et vous procurer un logiciel de communication.

Si vous disposez d'un terminal vous permettant l'accès à une machine Unix, qui, elle, est sur le réseau, vous pouvez accéder à travers cette machine à tous ces logiciels, mais au moyen d'une interface moins conviviale.

Un bibliothécaire isolé, qui n'a que des accès *via* Transpac pour interroger des serveurs documentaires tels Dialog ou l'ESA, et qui veut accéder à Internet pour ses ressources documentaires, doit s'adresser à un centre de calcul connecté à Internet. Dans ce cas, il utilise Transpac.

Les possibilités documentaires sur Internet se caractérisent par leur ampleur, leur diversité, la souplesse des nouveaux systèmes d'information qui se mettent aujourd'hui en place pour les gérer.

Internet est un puissant outil de transformation de la politique documentaire. Il met à disposition de tous l'information rétrospective comme l'information instantanée, – information dont nous avons donné un aperçu lorsque nous avons abordé la question des *newsgroups*. Par la messagerie, les connexions lointaines, par les possibilités de transfert de fichiers, Internet encourage la collaboration « distribuée » ; ainsi, Internet permet actuellement la constitution de serveurs locaux répartis sur ce réseau donnant à chaque laboratoire la possibilité d'exposer sur le réseau l'ensemble de ses publications, de les offrir à la communauté.

Le réseau national Renater, infrastructure de réseaux régionaux, permet l'accès au réseau Internet. L'année 1993 devrait en voir l'extension à tout le territoire et offrir ainsi de larges possibilités aux bibliothèques universitaires.

Décembre 1992

BIBLIOGRAPHIE

1. Archimbaud, Jean-Luc, « Eudora, une messagerie gratuite », *Le Microbulletin*, n° 44, avril-mai 1992, p. 109.

2. Archimbaud, Jean-Luc, « FTP anonymous », *Le Microbulletin*, n° 45, juin-sept. 1992, pp. 99-104. Et *Machine urec.urec.fr* (134.157.44), fichier */pub/ip/microbulletin/ftp.anonymous*.

3. Barthelemy, Pierre, « Le projet Wais », à paraître, *Le Microbulletin*. *Machine urec.urec.fr* (134.157.44), fichier */pub/ip/microbulletin/liste.bibs*.

4. Barthelemy, Pierre, « L'accès aux bibliothèques à travers l'Internet », *Le Microbulletin*, n° 47, déc.-janv. 1992, pp. 160-165. Et *Machine urec.urec.fr* (134.157.44), fichier */pub/ip/microbulletin/WAIS*.

5. Bessières, Henri, « Internet : la fièvre gagne les entreprises », *Telecoms magazine*, n° 20, mars 1993, p. 22-25.

6. Comer, Douglas, « TCP/IP : architecture, protocoles, applications, Paris, InterEditions, 1992, 572 p.

7. Farley, Laine, (ed.) « Strategies for selection and use », *machine ftp.imag.fr* (129.88.32.1), fichier */pub/DOC/INTERNET/libcat.gui*, août 1991.

8. Frey, Donnaly, Adams, Rick, *The directory of electronic mail : addressing and networks*, 2^e éd., A Nutshell Handbook Ser. O'Reilly & Ass., 1989, 308 p..

9. Kohoe, Brendan P., « Zen and the art of the Internet : a beginner's guide to the Internet », 1^{re} éd., January, 1992, 94 p. *Machine ftp.imag.fr* (129.88.32.1), fichier *postscript /pub/DOC/INTERNET/zen-1.0.ps.Z*.

10. Ed Krol, *The Whole Internet User's Guide*, O'Reilly, Sebastopol, Ca, 1992, 376 p.

11. Laquey, Tracy, Ryer, Jeanne C., « The Internet Companion : a beginner's guide to global networks », Addison-Wesley, 1992. Ed. électronique des 2 premiers chapitres, *Machine world.std.com* (192.74.137.5), fichier */OBS/The.Internet.Companion/The.Internet.Companion*.

12. Laquey, Tracy, *The user's directory of computer networks*, Bedford, MA : Digital Press, c. 1990, 630 p.

13. Michau, Christian, « A propos du réseau Renater », *Le Microbulletin*, n° 44, mai 1992, p. 117-121.

14. Ricou, Olivier, « Le Re'Zo : histoire et guide pour le novice », oct. 1992. *Machine ftp.imag.fr* (129.88.32.1), fichier *postscript /pub/DOC/INTERNET/Le.Rezo.ps*.

12. Hans NUSSBACHER, « Comparaison between Gopher, WAIS and WWW », *comp.infosystems.wais.*, 1^{er} oct. 1992.