

La recherche d'informations sur l'Internet

Répertoires et moteurs de recherche

par Pierre-Yves Duchemin

Bibliothèque nationale de France

Le réseau Internet est utilisé dans la communauté scientifique et documentaire comme un nouveau et puissant moyen de communication. Grâce aux protocoles TCP/IP, Internet permet l'interconnexion d'un grand nombre de réseaux entre un grand nombre de pays. L'Internet (International network), en français « l'interréseau », est un réseau de réseaux qui peuvent être aussi bien des réseaux locaux que des réseaux à longue distance. La participation de ces réseaux constitue ainsi un seul réseau virtuel à la disposition de l'utilisateur.

À l'origine, Internet permettait d'échanger du courrier *via* la messagerie électronique, de participer à des débats sur les sujets les plus divers et les plus techniques grâce aux listes de discussion et de transférer des fichiers. Très tôt, il a servi de support à l'interrogation de catalogues de bibliothèques ou de banques de données documentaires. À côté de ces premiers services, de nouveaux services facilitant la coopération et l'accès à l'information sur réseau sont apparus, notamment les serveurs FTP, Telnet, Gopher, WAIS et, plus récemment, World Wide Web. Ce dernier système a été un immense progrès, non seulement en gérant le texte, l'image fixe ou animée et le son, mais aussi en intégrant tous les autres serveurs d'information et en rendant « transparentes » leurs particularités pour l'utilisateur. Les serveurs disponibles sur l'Internet sont très nombreux mais l'information qu'on y trouve, si elle est très variée et en très grande quantité, est malheureusement parfois de qualité inégale.

Pour « séparer le bon grain de l'ivraie », de nouveaux services de recherche et de découverte d'information deviennent indispensables.

Le fonctionnement

Des millions d'ordinateurs de technologies différentes, fonctionnant sur des systèmes d'exploitation différents, utilisant des logiciels différents, etc., réussissent à communiquer entre eux grâce au protocole TCP/IP.

Les protocoles TCP/IP

Le protocole TCP/IP (Transfer control protocol/Internet protocol) regroupe l'ensemble des protocoles de télécommunications qui gère la transmission des données en permettant l'interconnexion de plusieurs systèmes sur l'Internet. Il se compose de deux grands types de normes : les normes TCP et les normes IP. Très schématiquement, les normes TCP découpent les données à transmettre en paquets, chaque paquet étant identifié par l'adresse de l'expéditeur et par la liste des données qu'il contient. Les normes IP ajoutent l'adresse du destinataire. Pendant la transmission, des routeurs orientent les paquets sur l'itinéraire le mieux adapté. À l'arrivée des données, l'ordinateur récepteur vérifie l'adresse IP, le contenu des paquets TCP et reconstitue l'ensemble des données.

URL

Une adresse URL (Uniform Resource Locator) est l'adresse d'une application disponible sur l'Internet, par exemple <http://www.bnf.fr>. La première partie de l'URL identifie le protocole de communication : `http` : hypertext transfer protocol ; ce pourrait être également Gopher ou Telnet, etc. La seconde partie de l'URL identifie le serveur `www.bnf.fr` : dans l'ordre inverse, serveur français (`fr`) de la Bibliothèque nationale de France (`bnf`) sur le Web (`www`).

L'URL ne désigne pas seulement un site Internet sur lequel on se trouve ; elle peut également donner les coordonnées électroniques d'un fichier en comportant des répertoires, sous-répertoires et fichiers. Il est ainsi possible d'accéder directement à un fichier précis : par exemple, les URL <http://www.bnf.fr/web-bnf/catalog/opaline.htm> et <http://www.bnf.fr/enluminures/accueil.htm> sont respectivement les adresses de la page de connexion à BN-OPALINE et la page d'accueil du serveur des 1000 enluminures « Charles V et son temps ».

L'URL est l'élément indispensable qu'il faut connaître... ou réussir à trouver pour accéder à une information sur l'Internet. À noter qu'une URL répond à une syntaxe extrêmement rigoureuse et que, sous peine d'interrogation infructueuse, il est absolument indispensable de respecter majuscules, minuscules, caractères spéciaux et caractères de ponctuation.

Les ressources de l'Internet

L'Internet offre trois grandes familles d'applications : le courrier électronique, la connexion à distance et le téléchargement de fichiers.

Le courrier électronique

Le courrier électronique (e-mail) constitue l'application la plus répandue sur l'Internet. Il permet de communiquer avec n'importe quel autre utilisateur dans le monde dont on connaît l'adresse électronique. Le principe en est simple : une boîte aux lettres électronique occupe un espace mémoire sur le disque dur de l'ordinateur-hôte dédié à la messagerie électronique ; tout nouveau message reçu est stocké dans cette boîte. Le courrier électronique permet d'envoyer des messages à un ou plusieurs utilisateurs ; il permet d'attacher au message un document saisi sur traitement de texte.

Les listes de messagerie

Il est également possible de s'abonner à une liste de messagerie. L'intérêt de cette fonctionnalité est d'être en contact avec des groupes de personnes qui ne parlent pas forcément la même langue, mais qui, toutes, partagent un intérêt commun. Les abonnements aux listes permettent de recevoir une documentation toujours à jour sur un sujet donné.

Les nouvelles

Il existe une autre façon de communiquer sur l'Internet avec des interlocuteurs du monde entier, dans n'importe quelle langue et sur n'importe quel sujet. Les groupes de nouvelles (*News*) permettent également de partager un intérêt commun avec de nombreux utilisateurs. À l'inverse des listes de messagerie, les messages ne sont pas versés automatiquement dans la boîte à lettres électronique et impliquent l'utilisation d'un programme de lecture qui permet la sélection et la sauvegarde des messages intéressants. La plupart des comptes Internet donnent accès à un programme de lecture de nouvelles. Comme dans une messagerie, chaque message est identifié par son expéditeur, son thème et sa date de rédaction.

La connexion à distance

La connexion à distance d'un ordinateur vers un autre est possible sur l'Internet grâce à divers outils.

Le téléchargement de fichiers

L'Internet permet l'accès à des informations textuelles, sonores, ou graphiques composées d'images fixes ou animées. Il existe également beaucoup de ressources gratuites (*freewares* ou *sharewares*) qui peuvent être des programmes, des logiciels ou des fichiers. Il est possible, grâce à un protocole FTP (file transfer protocol), d'envoyer ou de recevoir des fichiers stockés dans un espace disque protégé et géré indépendamment du système d'exploitation du serveur pour garantir la sécurité d'accès. Une bibliothèque est ainsi en mesure de récupérer des données bibliographiques d'un site distant.

Certains ordinateurs disposent du programme de transfert de fichiers FTP. La procédure FTP normale consiste ainsi à utiliser le nom d'utilisateur et un mot de passe. Elle permet de butiner dans un répertoire et de sélectionner les fichiers à transférer, mais ne permet pas de les lire. Le protocole FTP ne permet pas l'utilisation du menu d'interrogation de la base de données, ni l'envoi de messages électroniques. La consultation ne peut s'effectuer qu'après transfert du fichier sur l'ordinateur destinataire.

Une autre procédure, « Anonymous FTP » (FTP anonyme) permet de se connecter comme visiteur anonyme à un serveur pour transférer un fichier ou un groupe de fichiers vers son ordinateur. Pour être accessibles à tous les utilisateurs du réseau, les fichiers sont placés dans un espace disque pouvant être lu par tous. Cette procédure anonyme ne nécessite pas de posséder un compte Internet. Néanmoins, l'utilisateur doit entrer son adresse électronique en guise de mot de passe. L'administrateur du service peut ainsi savoir qui consulte sa base de données. Cette procédure est très couramment utilisée pour télécharger des fichiers.

La procédure FTP est très riche et fourmille d'informations utiles et variées. Son inconvénient majeur est qu'il faut d'abord savoir sur quel serveur chercher pour

trouver une information. Ensuite, une fois le serveur localisé, il est nécessaire de le parcourir par sa structure arborescente, ce qui peut parfois prendre du temps.

Accès aux services de recherche d'information sur l'Internet

Telnet

Telnet est un ensemble de programmes qui émule un micro-ordinateur en terminal d'un ordinateur-hôte. Il est ainsi possible d'ouvrir une session comme un utilisateur local et d'utiliser les applications et les programmes disponibles sur la machine-hôte. Un accès Telnet est prévu pour la seule consultation d'un catalogue et non pour la récupération de notices bibliographiques. Il s'effectue généralement par l'accès public en ligne (OPAC) : l'utilisateur distant interroge ainsi la base de données exactement sur la même configuration d'écran et avec le même menu que s'il se trouvait physiquement dans l'établissement. Cela suppose la connaissance préalable du menu d'interrogation de l'OPAC dans sa langue originale.

Une connexion *via* Telnet s'effectue grâce à un émulateur de type VT ; l'émulateur le plus courant est le VT 100, mais il existe également le VT 220 et le VT 320. Il est possible d'interroger un site Telnet par son adresse numérique (adresse IP) ou par son « mnémonique » : <telnet://opale02.bnf.fr> ou 193.50.133.201 (BN-OPALE), <telnet://opaline02.bnf.fr> ou 193.50.133.202 (BN-OPALINE).

<http://www.lights.com/hytelnet>
<http://moondog.usask.ca/hytelnet>
<http://library.usask.ca/hytelnet/>

Hytelnet est le répertoire international des catalogues de bibliothèques accessibles selon Telnet. Sans prétendre à une exhaustivité complète, c'est de loin l'outil le plus puissant disponible pour accéder à des données bibliographiques.

Gopher

À la différence de Telnet qui nécessite au préalable de connaître l'adresse électronique de l'ordinateur sur lequel on veut se connecter, Gopher est un programme

de balayage qui intègre les adresses des différents sites Internet. Gopher est ainsi très facile d'utilisation et beaucoup plus convivial que Telnet qui oblige à utiliser des lignes de commande Unix.

Gopher est à la fois un système d'information et un système de navigation sur réseau. Comme système d'information, la structure de Gopher est arborescente et fonctionne à l'aide de menus ; l'utilisateur dispose de plusieurs commandes pour explorer les fichiers. Comme système de navigation, Gopher organise les serveurs d'information pour faciliter la navigation et la découverte d'informations. Le protocole utilisé est propre à Gopher. Bien que moins utilisé depuis le développement du Web, Gopher reste un serveur très répandu sur l'Internet grâce à sa souplesse d'utilisation et à son interface qui ne nécessite aucune interface graphique : un simple terminal télétype suffit pour consulter un serveur Gopher. L'inconvénient majeur de Gopher est sa structure fixe et arborescente qui peut être rapidement parcourue si l'application consultée est de petite taille ; dans le cas contraire l'accès à l'information pertinente n'est pas immédiatement garanti.

WAIS, Archie, Veronica

WAIS, Archie et Veronica sont des programmes de recherche qui permettent à l'utilisateur d'accéder à des ressources sur l'Internet. Ce sont des outils très utiles quand on sait ce que l'on cherche, mais qu'on ne sait pas où le trouver. L'inconvénient est qu'aucun d'eux ne couvrant l'ensemble des applications Internet disponibles, il faut généralement les interroger tous les trois.

WAIS

WAIS (Wide area information server) existe depuis 1991 et est une des premières applications de la norme Z39.50.

C'est un logiciel du domaine public qui fonctionne selon le modèle client/serveur. WAIS est un serveur d'information conçu pour faciliter la recherche de documents. Il indexe tous les termes des documents recensés pour créer un fichier inverse (index). C'est un outil de recherches utilisé pour effectuer des requêtes sur des index en texte intégral. L'utilisateur doit sélectionner un index et définir en langage dit « naturel » le ou les mots sur lesquels il désire effectuer sa recherche ; en réponse, WAIS affiche la liste des documents qui, dans l'index sélectionné, contiennent les mots recherchés. L'inconvénient majeur de WAIS est de ne pas fournir une description correcte d'une base et de ne proposer qu'un fichier source très succinct créé par l'administrateur de la base.

Archie

Archie est un outil, éventuellement accessible par Telnet, qui permet de retrouver des informations parmi l'ensemble des nombreux fichiers FTP d'archives. Archie indexe périodiquement les fichiers de tous les sites concernés. L'utilisateur doit définir l'objet de sa requête par des mots en anglais, qu'il est possible de tronquer. Les fichiers sélectionnés par Archie peuvent, le cas échéant, être téléchargés sur l'ordinateur du demandeur. Archie est un service très consulté car il permet de trouver rapidement l'information utile mais son inconvénient majeur est son indexation des noms de fichiers souvent insuffisante.

Veronica

Veronica (Very easy rodent oriented network information for computer access) a été créé en 1993. C'est un outil de recherche très utile sur Gopher : Veronica aide à la localisation de serveurs contenant une information précise, grâce à un index obtenu à partir des menus Gopher.

L'utilisateur doit définir en anglais l'objet de sa requête. Veronica propose alors un menu Gopher qui rassemble tous les menus Gopher qui répondent à la recherche. La consultation se poursuit alors comme à partir de n'importe quelle application Gopher. L'utilisateur doit veiller à fournir à Veronica une recherche aussi précise que possible car la puissance de cet outil, qui parcourt tous les serveurs Gopher de la planète, peut générer beaucoup de « bruit ». L'inconvénient majeur de Veronica est de n'indexer que les titres, ce qui donne des possibilités de recherche assez réduites.

World Wide Web

Le World Wide Web est l'un des plus puissants modes de navigation sur l'Internet. À la différence de Gopher qui permet une navigation par menus hiérarchiques et arborescents, le Web est construit sur la notion d'hypertexte dans une architecture client/serveur. La navigation par liens hypertextuels permet la consultation de services toujours plus nombreux, parmi lesquels des catalogues de bibliothèques ou de centres documentaires, mais également toutes sortes d'applications multimédias : on y trouve en effet des bases textuelles, des bases graphiques d'images fixes ou animées, éventuellement accompagnées de bandes sonores. Cette navigation est rendue possible par l'utilisation d'un « navigateur » (*brouser* ou « fureteur »), logiciel d'application qui permet de se déplacer dans l'Internet.

Le principe de base du Web est celui du document hypertexte réparti. L'hypertexte consiste à coder de manière distinctive un mot ou un groupe de mots dans un texte. Ce codage, totalement transparent pour l'utilisateur, lie ce mot ou ce groupe de mots à un autre texte ou à un autre fichier, voire à une autre base de données sur un autre site. Le codage utilisé est un mot souligné. Au passage sur un lien

« Annuaire » électroniques

FOUR11

<http://www.four11.com/>

Presqu'un million d'adresses électroniques recensées. Troncature non autorisée. Four11 retrouve une adresse e-mail à partir d'un nom et d'un prénom d'utilisateur.

INTERNET ADDRESS FINDER

<http://www.iaf.net/>

Plusieurs millions d'adresses électroniques recensées. Troncatures droites et gauches autorisées.

OKRA

<http://okra.ucr.edu/okra/>

Cet annuaire recense 5,5 millions d'adresses électroniques. Troncature non autorisée.

hypertexte, la flèche, qui matérialise habituellement le curseur, se transforme en une main ; il suffit alors de cliquer sur la souris pour établir le lien hypertexte.

Les applications disponibles sur le Web sont généralement d'une présentation beaucoup plus conviviale et agréable que les applications Gopher ou Telnet : elles

présentent une image graphique, grâce à l'utilisation de couleurs, de polices de caractères différentes et d'une mise en page, parfois animée par des objets « Java ». Cet aspect graphique est obtenu par une structuration des fichiers électroniques en format HTML (hypertext markup language), sous-partie de la norme SGML. Les communications sur le World

Wide Web sont assurées par le protocole HTTP (hypertext transfer protocol). HTTP offre en outre l'accès à d'autres applications, parmi lesquelles Gopher, Telnet, WAIS et des serveurs FTP. Cette intégration des autres serveurs d'information est l'un des grands avantages du Web, puisqu'elle lui permet de masquer les particularités des différents protocoles d'accès.

Guides thématiques

BUBL

<http://www.bubl.bath.ac.uk/BUBL/>

Site du Bulletin Board for Libraries. Permet une recherche thématique alphabétique ou selon les indices de la Classification décimale universelle (CDU).

CARREFOUR

<http://carreWfour.net>

Recherche thématique uniquement sur des sites francophones.

CLEARINGHOUSE

<http://www.clearinghouse.net/>

Existe depuis la mi-1993. Recherche thématique selon douze critères et opérateurs booléens de guides et de pages Web.

CYBERSTACKS

<http://www.public.iastate.edu/~CYBERS-TACKS/>

Applications scientifiques et technologiques, indexées selon la Classification de la Bibliothèque du Congrès (LCSH). [- s'obtient avec la combinaison Alt + 0126]

DEJANEWS

<http://www.dejanews.com/>

Service très efficace de recherche sur les listes de nouveautés : la recherche s'effectue par date, nom de *newsgroup* ou auteur.

EUREKA

<http://www.eureka-fr.com>

Recherche thématique sur sites francophones.

FRANCITÉ

<http://www.i3d.qc.ca/francite.htm>

Existe depuis janvier 1996. Recherche thématique selon quatorze catégories.

GALAXY

<http://www.einet/galaxy.html>

<http://galaxy.einet.net/>

Recherche thématique selon onze critères principaux de serveurs Web qui permet en outre la recherche en plein texte.

IMAGINET

<http://www.imagnet.fr/ime/noel.html>

Important point de départ pour les « internautes » débutants.

INFOMINE

<http://lib-www.ucr.edu/>

Répertoire thématique de serveurs Web.

MAGELLAN

<http://www.mckinley.com>

Existe depuis la mi-1995. Recherche thématique selon vingt critères principaux. Le grand intérêt de Magellan est l'évaluation du service (de 1 à 4 étoiles) par un véritable « comité de lecture ».

NOMADE

<http://www.nomade.fr>

Existe depuis février 1996. Répertoire de sites Web francophones (France, Europe, Canada, etc.). Nomade décrit chaque site de manière claire et complète. Il propose deux modes de recherche par mots clés et par navigation dans une arborescence thématique.

THE VIRTUAL LIBRARY

<http://www.w3.org/hyoertext/DataSources/WWW/Servers.html>

<http://www.info.cern.ch/hypertext/DataSources/bySubject>

Répertoire de sites Web classés par pays. La Virtual Library permet également l'interrogation par critères thématiques.

UREC

<http://www.urec.fr/France/web.html>

Répertoire très efficace du CNRS, classé par thèmes. Une référence en ce qui concerne l'Internet français.

VIRTUAL TOURIST

<http://www.vtourist.com/webmap>

Répertoire de sites Web classés par pays ou par régions accessibles en cliquant sur une carte dont les zones géographiques sont en fait des liens hypertextes.

W3 SERVERS

<http://www.w3.org/hypertext/Datasources/WWW/Servers.html>

Recense des serveurs Web dans le monde, classés géographiquement puis alphabétiquement.

WEBCATS

<http://library.usask.ca/hywebcat>

WebCats recense les catalogues de bibliothèques accessibles sur le Web et présentant une interface d'interrogation HTML.

YAHOO

<http://yahoo.com>

<http://www.yahoo.com/>

<http://www.akebono.stanford.edu/yahoo/>

Yahoo (Yet Another Hierachically Organized Oracle) existe depuis 1994. Très puissant et très facile d'utilisation, c'est l'un des plus importants guides thématiques du Web. Il offre des répertoires thématiques spécialisés (Yahooligans pour les enfants, etc.) et permet la recherche par thème ou par mot clé.

<http://www.yahoo.fr>

Une version française existe depuis octobre 1996.

YANOFF

<http://www.spectracom.com/islist>

La liste de Scott Yanoff est l'un des plus anciens répertoires de recherche sur le Web. Il propose une recherche selon 92 critères thématiques.

YELLOW WEB

<http://www.yweb.com/home-fr.html>

Existe depuis février 1996. Propose une recherche thématique ou par mots clés selon un menu multilingue dont le français.

Outils automatiques de recherche

A2Z

<http://a2z.lycos.com/>

Répertoire du moteur de recherche Lycos.

ALIWEB

<http://www.cs.indiana.edu/aliweb/search>

Aliweb (Archie Like Indexing) est un service de recherche dont le fonctionnement sur le Web est identique à celui d'Archie pour les fichiers FTP.

ALTAVISTA

<http://www.altavista.com>

<http://altavista.digital.com/>

Disponible depuis décembre 1995, il est rapidement devenu le moteur de recherche « incontournable ». C'est l'un des moteurs de recherche les plus complets et les plus rapides. Il est vrai que sa puissance de recherche est phénoménale et qu'il permet de répondre à (presque) toutes les questions sans trop de « bruit ». Alta Vista recense environ 30 000 adresses, 3 millions d'adresses et 30 millions de pages Web. La plupart des références sont indexées en plein texte.

Version française de Altavista :

<http://altavista.telia.com/cgi-bin/telia?country=fr&lang=fr>

ECILA

<http://www.ceic.com/ecila/>

Très efficace service de recherche de sites Web situés en France. Il offre en outre une sélection des nouveautés sur l'Internet en France.

EXCITE

<http://www.excite.com/>

Existe depuis décembre 1995. Il recense environ 50 000 adresses et les références sont indexées en plein texte.

HOTBOT

<http://www.hotbot.com/>

Existe depuis mai 1996. Moteur très puissant qui utilise le moteur de recherche Inktomi également utilisé par l'université de Berkeley. Il recense environ 54 000 adresses et la plupart des références sont indexées en plein texte.

INFOSEEK

<http://www.infoseek.com/>

<http://www2.infoseek.com/>

<http://ultra.infoseek.com/>

Existe depuis le début de 1995. Il recense aujourd'hui plus de 50 000 adresses URL et les références sont indexées en plein texte. Ce service

fonctionne normalement sur abonnement mais il est possible de l'utiliser gratuitement dans une version « bridée » qui n'offre que dix réponses par requête.

INK TOMI

<http://inktomi.berkeley.edu/query.html>

Système de recherche puissant qui permet le croisement de dix critères. La fréquence de la mise à jour laisse malheureusement à désirer.

LYCOS

<http://www.lycos.com/>

<http://lycos.cs.cmu.edu/>

Apparu en mai 1994, Lycos est l'un des doyens des moteurs de recherche. Il recense environ 66 000 adresses, 37 millions de liens hypertextes et les références sont accessibles par un résumé indexé en plein texte.

OPEN TEXT

<http://www.opentext.com/>

Existe depuis le début de 1995. Ce moteur a eu son heure de gloire mais semble aujourd'hui tomber en désuétude. Il recense environ 20 000 adresses et les références sont indexées en plein texte.

THE GLOBAL NETWORK

<http://nearnet.gnn.com/gnn/GNNhome.html>

Puissant outil de recherche sur le Web, accessible sur plusieurs sites miroir.

WEB CRAWLER

<http://www.webcrawler.com/>

<http://webcrawler.cs.washington.edu>

<http://webcrawler.com/>

Existe depuis avril 1994. À l'origine, projet de l'université de Washington, il a été racheté par America Online en mars 1995, puis par Excite en novembre 1996. Il continue néanmoins à exister de façon autonome. Il recense environ 220 000 adresses et les références sont accessibles en plein texte.

WORLD WIDE WEB WORM

<http://wwwwww.cs.colorado.edu/wwwwww>

World wide web worm est un outil de recherche très puissant sur le Web, mais sa mise à jour ne semble plus assurée.

WOW IT'S BIG

<http://www.mit.edu:8001/people/mkgray/compre3.html>

Malgré sa taille, liste non exhaustive de serveurs Web, classés par pays et par domaine (.edu, .gov, .com, etc.).

Recherche d'information

Les serveurs d'information décrits ci-dessus constituent un ensemble de ressources gigantesque qu'il est souvent malaisé d'utiliser pleinement par manque d'outils adéquats de recherche et de découverte d'information : c'est une vérité de La Palice mais, quand on sait où chercher, on peut consulter un serveur et on

trouve l'information. L'inconvénient de cette approche est évidemment que la consultation ne peut être effectuée que si l'on sait à quel serveur s'adresser. Les systèmes d'information tels que Gopher ou Web permettent d'organiser les serveurs d'information pour faciliter la navigation et la recherche de l'information, mais il est en fait assez difficile de trouver rapidement ou de localiser à coup sûr l'informa-

tion : leur taille gigantesque est en effet un obstacle non négligeable.

C'est pourquoi on assiste au développement rapide de services qui permettent de trouver les serveurs répondant à la recherche effectuée. Ces services permettent de trouver plus facilement de l'information sur Internet en guidant l'utilisateur, en effectuant la recherche sur plusieurs serveurs, voire en proposant des choix.

Robots de recherche

ALL4ONE

<http://all4one.com/>

Robot qui utilise Altavista, Lycos, Yahoo et Webcrawler.

ALL IN ONE

<http://www.albany.net/allinone/>

<http://www.media-prisme.ca/allsrch.html> (version francophone)

Moteur de recherche dont l'interface d'interrogation « camoufle » plusieurs outils de recherche.

FIND-IT

<http://www.iTools.com/find-it/find-it.html>

Moteur de recherche qui utilise les services d'autres outils, notamment Altavista, Open Texte et Yahoo.

META CRAWLER

<http://metacrawler.cs.washington.edu:8080/index.html>

Robot qui utilise les services de Galaxy, Infoseek, Yahoo et Webcrawler.

PROFUSION

<http://www.designlab.ukans.edu/ProFusion.html>

Robot qui utilise les services de Altavista, Excite, Infoseek, Lycos, Open Text et Webcrawler.

SAVVYSEARCH

<http://guaraldi.cs.colostate.edu:2000/form>

<http://guaraldics.colostate.edu:2000/form?lang=french> (version française)

Robot qui utilise les services de 24 outils de recherche dont Aliweb, Altavista, Dejanews, Excite, Galaxy, Infoseek, Lycos, Magellan, OpenText, Webcrawler, Yahoo, etc. Disponible en 23 langues dont le français.

STARTING POINT

<http://www.stpt.com/>

Interface d'interrogation de seize outils dont Archie, Aliweb, Altavista, Excite, Infoseek, Inktomi, Lycos, Yahoo, Webcrawler, etc.

SURF CENTRAL

http://www.aztech-cs.com/aztech/surf_central.html

Interface d'interrogation de 14 outils dont Altavista, Excite, Infoseek, Inktomi, Lycos, Magellan, Open Text, Webcrawler, etc.

THE INTERNET SLEUTH

<http://www.isleuth.com/>

Interface d'interrogation d'une quinzaine de moteurs dont Altavista, Infoseek et Lycos.

USE IT !

<http://www.he.net/~kamus/usefr.htm>

Robot d'origine italienne qui utilise ensemble ou un par un les services d'une quinzaine d'outils de recherche.

W3 SEARCH ENGINES

<http://137.142.42.95/W3SearchEngines.html#Tools>

Interface qui permet, grâce à un menu, la connexion directe à différents moteurs de recherche.

Il existe deux grandes catégories d'outils de recherche sur l'Internet : les répertoires et les moteurs de recherche.

Les *répertoires* sont généralement réalisés par des équipes de chercheurs qui passent leurs journées à explorer le Web à la recherche de nouveaux sites.

Chaque nouveau site rencontré est analysé succinctement et, après examen, indexé selon une classification malheureusement trop souvent non normalisée. Il est possible de signaler l'existence d'un nouveau site à l'équipe chargée du répertoire. Yahoo est un exemple très connu de répertoire indexé.

Les *moteurs de recherche*, souvent baptisés *crawlers* ou *spiders*, sont généralement des ordinateurs qui explorent systématiquement le Web et qui créent aussi systématiquement des catalogues de pages Web. Les nouveaux sites rencontrés sont indexés sans contrôle du contenu. Alta Vista est un exemple très connu de moteur de recherche.

Bibliographie

- Andrieu Olivier, *Méthodes, outils de recherche sur l'Internet*, Paris, Eyrolles, 1996, 240 p.
- Aymonin David, «L'information scientifique et technique sur Internet. Ressources disponibles et méthodes de recherche», in *Les Nouvelles Technologies dans les bibliothèques*, sous la dir. de Michèle Rouhet, Paris, Éd. du Cercle de la Librairie, 1996, p. 241-276.
- Duda Andrzej et Perret Stéphane, «Services de recherche d'information sur Internet : accès à l'information répartie», in *Les Nouvelles Technologies dans les bibliothèques*, sous la dir. de Michèle Rouhet, Paris, Éd. du Cercle de la Librairie, 1996, p. 331-348.
- Lardy Jean-Pierre, *Recherche d'information dans Internet : outils et méthodes*, Paris, ADBS, Sciences de l'information. Etudes et techniques. 1996, 96 p.
- Le Moal Marie-Claude (dir.), *La Recherche de l'information sur les réseaux*, sous la dir. de Marie-Claude Le Moal et Bernard Hidoine, Paris, ADBS, Sciences de l'information. Études et techniques. 1996, 253 p.
- http://www-scd-ulp.u-strasbg.fr/urfist/Recherche_information/sommaire.htm
Chercher l'information : le défi, réalisé par l'URFIST de Strasbourg et régulièrement mis à jour.
- <http://wwwbibl.ulaval.ca/vitrine/giri>
GIRI (*Guide d'initiation à la recherche sur Internet*) et GIRI2 (*Guide des indispensables de la recherche sur Internet*) réalisés par le Groupe de travail sur l'accès aux ressources documentaires du CREPUQ (Université Laval, Québec).