

ENSSIB

Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques

Université

Claude Bernard - Lyon I

DESS en INFORMATIQUE DOCUMENTAIRE

Rapport de Stage

L'internet en Documentation Scientifique

Jérôme DEVIDAL

Sous la direction du Dr Thierry LECRIQUE

INSTITUT DE RECHERCHE PIERRE FABRE

1997

BIBLIOTHEQUE DE L'ENSSIB



8115153

ENSSIB

Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques

Université

Claude Bernard - Lyon I

DESS en INFORMATIQUE DOCUMENTAIRE

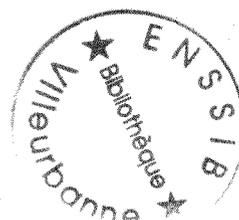
Rapport de Stage

L'internet en Documentation Scientifique

Jérôme DEVIDAL

Sous la direction du Dr Thierry LECRIQUE

INSTITUT DE RECHERCHE PIERRE FABRE



1997

1997
17 ST
07

L'Internet en Documentation Scientifique :

La documentation classique se trouve face à un nouvel outil qui doit faire partie de la palette de ses ressources. L'Internet constitue un nouveau canal pour véhiculer l'Information Scientifique et Technique. De la même façon que l'informatique s'est intégrée à côté de l'imprimerie classique, l'Internet se positionne entre les bases ASCII et les rayonnages des bibliothèques. Comme toutes les nouvelles solutions, il apporte des fonctions inusitées jusqu'alors.

Mots clés : Internet, Recherche, Industrie pharmaceutique, Base donnée
Donnée factuelle, Documentation, Méthode, Site web, Industrie chimique.

The use of Internet in Scientific Documentation :

Documentation is faced to a new computerized tool : Internet.

It is used as a new vehicle for scientific information retrieval. As computer has been integrated next to printing, Internet is standing between ASCII bases and bookcases.

Like each innovation, fonctions unused since now are offered by Internet.

Keywords : Internet, Research, Pharmaceutical industry, Database, Factual data,
Documentation, Methods, Web site, Chemical industry.

TABLE DES MATIERES

| | |
|--|-----------|
| TABLE DES MATIERES | 1 |
| HISTORIQUE DU GROUPE PIERRE FABRE | 3 |
| L'INTERNET MEDIA POLYMORPHE | 5 |
| CHAPITRE I :PRESENTATION DE L'INTERNET | 6 |
| 1. SERVICES PROPOSES SUR L'INTERNET | 6 |
| 1.1. HISTORIQUE | 6 |
| 1.2. SERVICES DE L'INTERNET | 7 |
| 1.3. ADRESSES ET URL | 8 |
| CHAPITRE II :L'INTERNET POUR LE CENTRE DE DOCUMENTATION SCIENTIFIQUE..... | 10 |
| 1. AVANTAGES POUR UNE DOCUMENTATION | 10 |
| 1.1. FOURNITURE DE DOCUMENTS VIA L'INTERNET | 10 |
| 1.2. FOURNITURE DE DOCUMENTS PRIMAIRES | 12 |
| 1.2.1. <i>Documents accessibles gratuitement</i> | 13 |
| 1.2.2. <i>Documents et services payants</i> | 14 |
| 1.3. ACCÈS À DE NOUVEAUX TYPES DE DONNÉES INEXPLOITÉES JUSQU'À CE JOUR | 15 |
| 1.3.1. <i>Cours et autres "education material"</i> | 15 |
| 1.3.2. <i>Serveur de prépublications</i> | 16 |
| 1.3.3. <i>Banques d'images</i> : | 16 |
| 1.3.4. <i>Groupes et listes de discussion</i> | 16 |
| 2. MISE EN ŒUVRE EN DOCUMENTATION SCIENTIFIQUE..... | 17 |
| 2.1. FORMATION DES FUTURS UTILISATEURS..... | 17 |
| 2.2. MAINTENANCE DU SERVICE..... | 17 |
| 2.2.1. <i>Support logiciel et matériel</i> | 17 |
| 2.2.1.1. Applications assistantes | 17 |
| 2.2.1.2. Logiciels utiles pour la gestion du poste | 19 |
| 2.2.1.2.1. Antivirus..... | 19 |
| 2.2.1.2.2. Logiciel de décompression..... | 19 |
| 2.2.1.2.3. Gestion des installations..... | 19 |
| 2.2.2. <i>Recherche d'Information Scientifique et Technique</i> | 20 |
| 2.2.2.1. Ressources d'informations disponibles | 20 |
| 2.2.2.1.1. Point d'entrée et carnet d'adresses | 20 |
| 2.2.2.1.2. Répertoires géographiques | 20 |
| 2.2.2.1.3. Répertoires thématiques | 21 |
| 2.2.2.1.4. Moteurs de recherche | 22 |
| 2.2.2.1.5. Guides de ressources | 23 |
| 2.2.2.1.6. Conférences électroniques..... | 24 |

| | |
|--|-----------|
| 2.2.2.1.7. Signalement des nouveautés | 26 |
| 2.2.2.2. Méthode générale | 26 |
| 2.2.2.3. Outils d'aide à la recherche | 28 |
| 2.2.2.3.1. Agents veilleurs [14] | 28 |
| 2.2.2.3.2. Agents aspirateurs ou Off-line browsers | 28 |
| 2.2.2.3.3. Agents de recherche intelligents | 30 |
| 2.2.2.3.4. Exemple d'aide à la recherche interne : <i>Excite for Web Servers</i> | 30 |
| CHAPITRE III : L'INTERNET POUR LE C.R.P.F..... | 31 |
| 1. L'INTERNET POUR LES "UTILISATEURS NON-DOCUMENTALISTES" | 31 |
| 2. POLITIQUE GLOBALE CHEZ PIERRE FABRE | 32 |
| 2.1. L'INTERNET AUJOURD'HUI | 32 |
| 2.1.1. <i>Courrier électronique au CRPF</i> | 32 |
| 2.1.1.1. Etude de l'existant | 32 |
| 2.1.1.2. Besoins | 32 |
| 2.1.1.3. Solutions à envisager..... | 33 |
| 2.1.1.3.1. Adresse unique | 33 |
| 2.1.1.3.2. Adresse partagée | 35 |
| 2.1.2. <i>Recherche sur le Web</i> | 35 |
| 2.2. SOLUTION FUTURE | 35 |
| 3. PRINCIPALES SOURCE EN CHIMIE ET BIOLOGIE SUR L'INTERNET .. | 36 |
| CHAPITRE IV : L'INTERNET EST UNE VITRINE | 41 |
| 1. DROIT ET SECURITE..... | 41 |
| 2. CONCEPTION..... | 42 |
| 3. EVOLUTION TECHNIQUE..... | 43 |
| CONCLUSION..... | 44 |
| ANNEXE : SIGNETS | 45 |
| BIBLIOGRAPHIE..... | 54 |



HISTORIQUE DU GROUPE PIERRE FABRE

- 1961** Mise au point par Pierre Fabre du premier produit d'origine phytochimique pour l'insuffisance veineuse : Cyclo 3[®].
- 1963** Acquisition des Laboratoires Inava.
- 1965** Acquisition des laboratoires Klorane et création de l'activité dermo-cosmétique.
- 1968** Création du Centre de Recherche Pierre Fabre à Castres.
- 1969** Acquisition des Laboratoires Ducray.
- 1970-1973** Implantation des premières filiales en Espagne, Allemagne et Italie.
- 1976** Prise de contrôle du premier fabricant d'édulcorant de synthèse : Edulcor.
- 1977-1978** Implantation aux Etats-Unis et au Bénélux.
Création des Laboratoires Galénic
- 1979** Acquisition de René Furterer
- 1981** Création de Shiseido France, joint-venture avec le deuxième groupe mondial de cosmétique.
- 1989** Création des Laboratoires Dermatologiques Avène. Mise sur le marché de Navelbine[®]. Acquisition des Laboratoires suisses Robapharm.
- 1990** Acquisition aux Etats-Unis de la société Physicians FormulaCosmetics.
Ouverture d'une station thermale à Avène. Création d'un centre de recherche en immunologie à Saint-Julien-en-Genevois. Création de D'Médica, société de distribution et de services pour la médicalisation à domicile.
- 1991** Création à Pau d'une unité spécialisée dans la production d'injectables et d'antibiotiques, agréée depuis aux normes FDA.
- 1992** Acquisition d'une usine de production à Gien. Création d'un nouveau centre de pharmacocinétique près de Castres.

- 1994** Création de l'Institut Klorane pour la protection et la bonne utilisation du patrimoine végétal. Obtention de l'AMM Navelbine® aux Etats-Unis.
- 1995** Dépôt du dossier d'AMM du Milnacipran (Ixel®) en France.
- Juillet 1997** Commercialisation de Ixel® en France.

Aujourd'hui, les Laboratoires Pierre Fabre constituent le 2^{ème} groupe pharmaceutique français indépendant et le 48^{ème} au classement mondial des sociétés pharmaceutiques en fonction du nombre de produits en recherche et développement (32 produits). L'entreprise emploie 6700 personnes dont 1570 à l'étranger et 5130 en France dont 3029 dans la région Midi-Pyrénées. En 1996, le Chiffre d'Affaire est estimé à 5.8 milliards de francs dont 30% provient des filiales internationales.



L'INTERNET MEDIA POLYMORPHE

Ce mémoire est le résultat du stage effectué dans le cadre du DESS Informatique Documentaire de L'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques de Lyon. Le stage s'est déroulé sur une période de 4 mois, au sein de l'équipe de Recherche Documentaire du Centre de Documentation Scientifique de l'**Institut de Recherche Pierre Fabre** à Castres (81).

Nous n'avons volontairement pas utilisé le terme de conclusion pour qualifier le travail décrit dans ce mémoire car s'il existe un domaine qui est en perpétuel mouvement (avec ses progrès et ses abandons), c'est bien celui de l'Internet.

Pour Robert Caillau, ingénieur du CERN dont les travaux sont à l'origine du protocole *hypertext*, "l'Internet" est désormais le terme consacré. Selon lui, « la France est le seul pays où l'on dit "sur Internet" au lieu de "sur l'Internet". On n'a donc toujours pas compris que l'Internet est une idée et non pas une marque de boisson américaine ». [1]

Ce rapport présente un peu de théorie, des réponses concrètes, des pistes pour aller plus loin. Cependant, la vitesse avec laquelle l'Internet évolue et révolutionne est tellement élevée que, si les solutions restent fondamentalement valables, leur mise en œuvre future sera inévitablement différente car l'outil aura été modifié.

Chapitre I :PRESENTATION DE L'INTERNET

L'Osid (observatoire des stratégies de l'information et de la documentation) a présenté au salon IDT 97, les résultats d'une étude analysant les influences de l'Internet sur les pratiques documentaires. Un panel de 100 entités a été interviewé parmi lequel 85% des personnes interrogées, affirment se servir d'Internet pour effectuer des recherches, alors que moins de 68% utilisent déjà la messagerie.[2]

Les aspects positifs sont les suivants : ouverture à de nouvelles sources d'information et gain de temps. Les aspects négatifs : lenteur de la transmission, complexité de la recherche et manque de formation.

Nous rajouterons ici le point le plus problématique des informations trouvées sur le Web : *l'absence de validité des documents gratuits.*

1. SERVICES PROPOSES SUR L'INTERNET

1.1. Historique

L'Internet est né au début des années 70 aux Etats-Unis quand l'Advanced Research Project Agency (Centre de projet pour la recherche avancée appartenant au département de la défense américain) décida de relier ses principaux centres de recherche pour partager les équipements informatiques, échanger des données et du courrier. Le but était de concevoir un réseau qui résiste à des attaques militaires telles que des bombardements tout en restant opérationnel ; Ainsi, il ne devait pas y avoir de point névralgique dans le réseau dont l'arrêt aurait provoqué le blocage de celui-ci. Les données devaient pouvoir prendre automatiquement un chemin différent en cas de coupure de liaison [1]. Ce système de réseau devait comprendre un *reroutage dynamique* c'est à dire que si l'une des liaisons du réseau est indisponible, les données circulant sont automatiquement dirigées vers d'autres liaisons.

A partir de ce système, les Etats-Unis avaient la possibilité de développer un réseau opérationnel (l'ancêtre de l'Internet actuel) et les utilisateurs, chercheurs et enseignants, qui y avaient accès devinrent rapidement enthousiastes.

1.2. Services de L'Internet

Les services correspondent aux "couches" successives qui se sont accumulées depuis la naissance de l'Internet.

Il y a Telnet pour transformer sa machine en terminal capable de se connecter sur de gros ordinateurs et aussi FTP pour récupérer et transférer tous les types de fichiers.

La messagerie électronique, le fameux e-mail, permet à tout le monde de communiquer par courrier [4]. Elle n'apporte que peu de nouveauté par rapport aux nombreux autres médias déjà existants. Elle est cependant le seul outil à réunir autant de qualités :

- **rapidité** : le temps de transit d'un message d'un utilisateur vers la boîte aux lettres de son correspondant est semblable au temps d'établissement d'une communication téléphonique.
- **asynchronisme** : contrairement au téléphone, il n'est pas nécessaire que les correspondants, soient disponibles en même temps pour être mis en rapport. Quel que soit le moment où l'émetteur envoie son message, le récepteur a toute latitude pour choisir le moment où il va le lire. Cette propriété est essentielle, dans les communications intercontinentales avec de grands décalages horaires, dans les communications entre de nombreuses personnes, mais aussi dans les communications de personne à personne pour lesquelles le téléphone et l'intrusion de sa sonnerie impérieuse constitue une gêne.
- **économie** : dans le cas de sites ayant payé leur connectivité IP globale, il n'y a aucun coût additionnel pour l'échange de messages.
- **données exploitables** : contrairement à un texte reçu par téléphonie, un texte reçu par messagerie peut être annoté, corrigé, ou réédité avant d'être éventuellement réexpédié.
- **communication de groupe** : un message peut être explicitement adressé à plusieurs personnes. Il existe en outre le mécanisme des listes de diffusion qui permet de s'adresser à une ou plusieurs communautés regroupées par thème de discussion.[1]

Les listes de discussion ou Listservs de USENET avec ses milliers de newsgroups ou forums de discussion permettent à chacun de s'exprimer librement et publiquement.

Gopher est apparu avec ses menus, puis plus récemment le World Wide Web (ou WWW ou W3 ou Web) basé sur l'hypertexte qui permettent de naviguer dans le réseau à la recherche d'informations.

WAIS, plus spécifique, est un ensemble de logiciels permettant de constituer et d'interroger des bases de données informatiques. Il s'agit plutôt d'un outil que d'un service.

En 1996, le Web a tout absorbé : il suffit de naviguer avec son logiciel Netscape ou Internet Explorer (il en existe d'autres) pour accéder aux ressources FTP, Gopher, WAIS, News, messagerie, etc...

1.3. Adresses et URL

Sur l'Internet, nous trouvons trois types d'adresses qui sont celles des machines, des personnes et des documents.

| <i>Objet</i> | <i>Adresse</i> | <i>Dénomination</i> |
|--------------|--|--|
| machine | 130.79.200.1 isis.u-strasbg.fr | adresse IP correspondant à un nom ou DNS |
| personne | nom@isis.u-strasbg.fr | e-mail ou adresse électronique |
| document | http://isis.u-strasbg.fr/pub/journal.html telnet://opale02.bnf.fr | Uniform Resource Locator (URL) |

Tableau 1 [4]

Dans le carnet d'adresses ou bookmark, nous notons les URL (Uniform Resource Locator) ou "manière uniforme de localiser des ressources" qui sont du type :

spécificateur_d'accès://nom_de_machine/répertoire/nom_de_fichier.extension

- ✦ *spécificateur_d'accès* : le sigle précédant les partie adresse et fichier d'une URL, concerne le type de protocole utilisé : *http* est le plus courant et indique que la page est un document Web auquel nous accédons grâce au protocole HTTP (HyperText Transfer Protocol). A la place de *http*, nous pouvons trouver d'autres termes qui indiquent au navigateur quels serveurs il doit appeler et comment le faire.
 - ✦ *ftp* : signifie que nous souhaitons accéder à un serveur de fichier à l'aide de FTP (File Transfer Protocol).
 - ✦ *telnet* : accède à un serveur fonctionnant en mode texte.
 - ✦ *gopher* : assure la compatibilité entre le réseau de documents gopher et le réseau Web.

- ◆ *news* : accède aux newsgroups.
- ▲ *nom_de_machine* : l'emplacement de la machine peut être saisie indifféremment en majuscules ou en minuscules.
- ▲ *répertoire/nom_de_fichier.extension* : la partie fichier de l'URL correspondant à la place du document sur le serveur, doit respecter la casse (majuscule ou minuscule).

Chapitre II : L'INTERNET POUR LE CENTRE DE DOCUMENTATION SCIENTIFIQUE

1. AVANTAGES POUR UNE DOCUMENTATION

L'Internet est né de la volonté des communautés scientifiques et universitaires de se doter d'un moyen de communication utilisable au niveau mondial. Les services qui se sont développés sur le réseau des réseaux sont donc fortement orientés vers un public scientifique et l'Information Scientifique et Technique y est très présente.

Cependant, l'état actuel de la technique et l'histoire même du développement du réseau font que l'information proposée est différente de celle que l'on trouve sur d'autres supports électroniques ou même sur papier : les procédures de validation de l'information n'existent pratiquement pas et la diffusion commerciale d'une information à haute valeur ajoutée commence seulement.

De plus le foisonnement d'informations sous des formes variées ainsi que la relative complexité des outils de navigation et de recherche rendent l'offre d'information confuse. En contrepartie, la possibilité de dialoguer avec d'autres utilisateurs transforme les modes de recherche et permet de passer de la recherche documentaire à la recherche d'informations au sens large, en contribuant aussi à la diffusion plus rapide des connaissances.

En 1997, le plus grand réseau informatique du monde entre dans sa phase de maturité : les développements en cours, la structuration des services, l'amélioration des outils logiciels permettent déjà de mener des recherches d'information avec méthode et efficacité, faisant de l'Internet une source indispensable d'information scientifique et technique.[5]

1.1. Fourniture de documents via l'Internet

L'IST (Information Scientifique et Technique) au sens "classique" est l'information contenue dans les articles, actes de congrès, thèses, ouvrages, brevets et rapports servant à la diffusion des connaissances. Dans ce sens elle est donc structurée et en général identifiable au travers des catalogues de bibliothèques ou des bases de

données bibliographiques et en texte intégral qui recensent la littérature scientifique mondiale.

Cependant une grande partie de la communication scientifique échappe à ce mode écrit pour se dérouler entre les chercheurs eux-mêmes: on parle alors du "collège invisible", ce *réseau informel* fait de rencontres, de dialogues, de courriers entre spécialistes d'un même sujet [5]. Un membre du "collège" peut s'adresser au groupe pour lui communiquer une information ou bien lui poser une question et en attendre des réponses pertinentes. Or, jusqu'à l'émergence des réseaux informatiques pour la recherche, il était difficile de participer à un "collège" sans être soi-même un spécialiste du domaine étudié. Avant la généralisation de l'Internet - rappelons que près de 160 pays sont aujourd'hui reliés par ce réseau - une recherche d'information scientifique et technique se faisait nécessairement par des canaux distincts et à l'aide d'outils différents : par les contacts personnels (téléphone, fax, rencontres lors des congrès), par l'exploration des bases de données (terminal, minitel ou micro-ordinateur) et bien sûr par la lecture de publications scientifiques.

Aujourd'hui l'Internet rend possible toutes ces recherches à l'aide d'un seul outil : l'ordinateur.

Il ne faut pas espérer trouver absolument tout sur l'Internet. Ce n'est en fait qu'une infrastructure de télécommunication (un réseau de distribution comme le réseau téléphonique). Les messages et les services qu'il transporte sont le fruit d'un travail humain réalisé pour répondre aux besoins d'autres humains à un moment donné, en fonction de possibilités techniques et technologiques propres à l'époque. Or il est évident aujourd'hui que la production d'Information scientifique et technique est une activité économique à part entière qui obéit à des impératifs de rentabilité. L'IST est coûteuse à produire et pour être rentabilisée, il faut la vendre à ceux qui en ont besoin et qui peuvent la payer. Certes, le réseau électronique facilite sa diffusion et permet d'abaisser les coûts d'édition, mais sans réduire les coûts de création. De plus il demande que le client dispose d'un équipement adéquat pour accéder à l'information ; il ne peut donc être utilisé par tous ceux qui ont besoin d'information, s'ils n'ont pas assez d'argent, d'équipement informatique. Sans compter qu'il n'est pas dit que tout le monde soit intéressé par cette forme (lire un livre reste plus agréable que lire un écran). Le papier reste le support idéal pour diffuser auprès du plus large public les connaissances

générales constituant le savoir commun de l'humanité (oeuvres littéraires, manuels).et les supports optiques comme le CD-ROM permettent la diffusion de très gros volumes d'information de façon économique et rapide.

Parallèlement, l'Internet est et restera pour un certain temps encore le support de l'information d'actualité permettant la diffusion rapide de nouvelles connaissances vers ceux qui peuvent et veulent en payer le prix.

1.2. Fourniture de documents primaires

Les informations que nous pouvons trouver sur l'Internet sont de trois types :

1. informations isolée : documents, images, sons, programmes.
2. informations circulantes :
 messages échangés dans les NEWS de Usenet
 messages échangés dans les listes de discussion.
3. bases de données :
 bibliographique, factuelle.

Pratiquement tous ces éléments d'information sont disponibles gratuitement à la consultation. Ils sont présentés par des individus ou des organisations qui en sont propriétaires [6]. Il existe cependant des services payants dont l'accès est limité aux seuls abonnés ou obligent au paiement en ligne à l'aide d'une carte de crédit. Parmi ces services, signalons notamment :

Les bases de données commerciales :

- exemples : MEDLINE, CHEMICAL ABSTRACTS, INSPEC, BIOLOGICAL ABSTRACTS.

Les serveurs et les producteurs des bases suivantes utilisent l'Internet pour informer leurs clients :

- Questel, Dialog, STN-CAS (serveur spécialisé en chimie), ISI (éditeur de Current Contents et Science Citation Index), INIST (producteur de la base PASCAL), BIOSIS (base Biological Abstracts).

Un grand nombre de périodiques scientifiques ou d'actualité :

- exemples : Journal of Biological Chemistry, Le Monde, Libération.

1.2.1. Documents accessibles gratuitement

La plupart des serveurs d'information qui composent Internet sont aujourd'hui accessibles librement. Ils permettent de lire et de récupérer gratuitement pour son propre usage :

- Des documents en texte intégral :
 - ⇒ articles scientifiques, notamment en physique et en mathématiques
 - ⇒ cours et manuels sur presque tous les sujets
 - ⇒ oeuvres littéraires du domaine public
 - ⇒ rapports d'études de grands organismes nationaux et internationaux
 - ⇒ traités internationaux, législation ;
 - ⇒ statistiques mondiales et nationales ;
 - ⇒ archives de groupes de discussion : bibliographies, notes d'information, les FAQ¹ (Frequently Asked Questions ou Foire Aux Questions) répondent aux questions que tout utilisateur d'Internet se posera un jour ou font le point sur un sujet intéressant les lecteurs d'un groupe. Ce sont en fait des guides réalisés par les animateurs de groupes de discussion. Les FAQ sont accessibles sur la plupart des serveurs universitaires.[5]
- Des documents techniques concernant les réseaux et l'informatique (manuels de logiciels, spécifications techniques). Ces documents sont de deux types :
 - ⇒ les RFC (Request for Comments)² décrivent le fonctionnement et les spécifications techniques de tout ce qui sert à faire fonctionner l'Internet : logiciels, matériel, protocoles.
 - ⇒ les FYI (For Your Information) sont des RFC pour débutants. La RFC 1392 par exemple est un glossaire des termes techniques d'Internet, la RFC 1462 a pour titre : "Qu'est-ce qu'Internet?".
- Des références bibliographiques :

Obtenues par l'interrogation des très nombreux catalogues de bibliothèques ou des quelques bases de sommaires de périodiques. Les éditeurs commerciaux

¹ Pour en savoir plus sur les FAQ et sur la manière d'y accéder : <http://www.best.com/~ii/internet/faqs.html>

² Pour en savoir plus sur les RFC et sur la manière d'y accéder : <http://www.nic.fr/Guides/>

présentent leurs catalogues en ligne et certains permettent même de commander leurs publications via Internet.

- Des données scientifiques :

Contenues dans des banques de données spécialisées : séquences nucléiques et protéiques, tableaux périodiques des éléments.

- Des images scientifiques (photographies, dessins, graphiques) :

Les problèmes de format et de taille se posent surtout avec ces documents.

1.2.2. Documents et services payants

Le succès du réseau Internet a amené les producteurs d'information commerciale à s'intéresser à ce nouveau média. De plus, la recherche de solutions techniques a permis de développer les transactions commerciales ou payantes sur Internet (cryptage et sécurisation des données confidentielles). D'ores et déjà, il est possible d'utiliser Internet pour accéder aux serveurs de banques de données bibliographiques commerciales mais sans avantage notable par rapport à un accès via les réseaux de télécommunication habituels (Transpac ou autre)[5]. En effet, la convivialité est améliorée, mais la puissance et la finesse des langages d'interrogation n'est jamais complètement transcrite. Les professionnels attendent beaucoup de la sécurisation des transactions pour permettre un accès généralisé via Internet.

Des services nouveaux apparaissent et sont beaucoup plus originaux. En voici quelques exemples :

- Revues de presses et journaux d'actualité personnalisés :

Dow Jones³ fournit un logiciel qui permet de composer des profils documentaires pour recevoir chaque jour les articles extraits de toutes les publications du groupe, selon des critères définis par l'abonné.

- Texte intégral d'articles, de périodiques, d'ouvrages tarifés à l'unité :

La lecture en ligne de documents devient une réalité. On peut déjà consulter dès sa parution le texte intégral du Journal of Biological Chemistry⁴, l'un des plus prestigieux périodiques en biologie.

³ Dow Jones : <http://bis.dowjones.com/>

⁴ Journal of Biological Chemistry : <http://www-jbc.stanford.edu>

- Bases de données spécialisées consultables avec des interfaces conviviales :
Sur abonnement forfaitaire ou au temps passé et au nombre de documents, il est possible d'accéder à des bases de données bibliographiques ou textuelles spécialement adaptées à l'Internet, c'est-à-dire bénéficiant d'interfaces graphiques très agréables grâce au mode client serveur. On peut citer les bases diffusées par Silver Platter⁵ selon l'ERL Technology", la Cambridge Structural Database⁶ ou le système SciFinder de Chemical Abstracts⁷ en chimie, les fichiers Dun & Bradstreet sur le serveur Prodigy.

1.3. Accès à de nouveaux types de données inexploitées jusqu'à ce jour

L'Internet est multimédia, il est possible de diffuser par le réseau :

- du texte
- du son
- de l'image fixe et animée
- des éléments informatiques : programmes et bases de données.

A ce titre Internet est à la fois un nouveau canal pour diffuser ce qui existait déjà, mais aussi un nouveau lieu de communication et de diffusion qui génère de nouvelles pratiques comme les groupes de News et les listes de discussion.

1.3.1. Cours et autres "education material"

➤ un article qui fait le point :

<http://www.elsevier.nl:80/section/chemical/trac/educate.htm>

➤ un serveur qui propose des exercices scientifiques en chimie et biologie : Science resource center : <http://www.lapeer.lib.mi.us/Chem/EastSciRes.html>

➤ World Lecture Hall recense les cours en chimie :

<http://www.utexas.edu/world/lecture/ch/>

➤ Guide to Chemistry Information : site original destiné aux étudiants, il présente les ressources selon le problème posé : <http://palimpsest.stanford.edu/cheminfo/>

➤ Martindale's Health Science Guide :

<http://www-sci.lib.uci.edu/~martindale/HSGuide.html>

⁵ Silverplatter et ERL : <http://www.silverplatter.com>

⁶ Cambridge Structural Database : <http://www.ccdc.cam.ac.uk/>

⁷ Scifinder from CAS : <http://www.cas.org/>

1.3.2. Serveur de prépublications

Les prépublications sont pour l'instant peu développées en chimie, nous ne trouvons que la chimie-physique : <http://www.chem.brown.edu/chem-ph.html> ainsi que des preprints en spectroscopie de masse des protéines : <http://128.122.10.5/preprints/contents.htm>

1.3.3. Banques d'images :

- Le projet *Visible Human* réalisée par la National Library of Medicine, est accessible en partie sur l'Internet avec des coupes, des reconstructions en 3D ainsi que des animations :
<http://www.uke.uni-hamburg.de/Institutes/IMDM/IDV/Gallery.html>
- Images d'histoire de la médecine : fournit un accès à quelques 60 000 images (reproductions photographiques, peintures) de la collection de l'US National Library of Medicine :
<http://oli.lhc.nlm.nih.gov/>
- Les acides aminés : <http://www.chemie.fu-berlin.de/chemistry/bio/amino-acids.html>

1.3.4. Groupes et listes de discussion

- Archive interrogeable par mots clés de tous les groupes et listes consacrés à la chimie, mis à jour toutes les 24 heures : Searchable Chemistry Mailing and Newsgroup Database : <http://merlin8.npac.syr.edu/chem/>
- Les groupes et listes de discussion en chimie sont assez peu nombreux et peu spécialisés (en comparaison de la biologie) :
<http://bionmr1.rug.ac.be/chemistry/overview.html>
- Les FAQ du groupe sci.chem font le point régulièrement sur les sources d'information en chimie, les débats d'actualité et les questions que se posent les chimistes du monde entier : à lire sur *Dejanews*⁸ en posant la question FAQ dans le groupe sci.chem : <http://www.dejanews.com/forms/dnq.html>



⁸ http://www.dejanews.com/home_ps.shtml

2. MISE EN ŒUVRE EN DOCUMENTATION SCIENTIFIQUE

2.1. Formation des futurs utilisateurs

Le premier mois du stage a permis la prise en main de l'outil en adéquation avec la compréhension des besoins. Le second mois a vu la mise en œuvre de la formation interne. En attendant que le projet de mise en place d'un intranet déploie un nombre suffisant de connexions, nous avons réalisé trois séances de présentation des possibilités de l'Internet. De plus, la personne chargée de transmettre les commandes à l'INIST a pu prendre en main ce nouvel outil après quelques tests d'essai du service.

Le poste que nous avons utilisé est destiné à devenir un libre service, notamment en ce qui concerne la messagerie électronique. De ce fait, la secrétaire scientifique du service de Zootechnie a été familiarisée avec la manipulation du système e-mail.

2.2. Maintenance du service

2.2.1. Support logiciel et matériel

Un micro-ordinateur équipé des logiciels adéquats, un modem ou une carte réseau et un accès à l'Internet suffisent pour retrouver des documents, interroger des bases de données bibliographiques ou factuelles et communiquer directement avec des groupes d'experts. Mais l'Internet n'est encore que le prototype des autoroutes de l'information annoncées par les médias ; comme tout prototype, il nécessite des réglages, des améliorations et surtout une simplification d'emploi pour être facilement utilisable par tous. L'intégration de tous les services et leur utilisation avec un seul logiciel est en route. Mais pour l'instant il faut apprendre à utiliser plusieurs logiciels pour exploiter les différents services proposés sur le réseau et disposer d'un lot de logiciels utilitaires pour manipuler, traiter et exploiter localement les documents récupérés (au sens large : textes, sons, images).

2.2.1.1. Applications assistantes

Aujourd'hui en 1997, en plus du navigateur il faut récupérer et installer tout un ensemble de logiciels supplémentaires pour profiter des capacités multimédia du Web.

| <i>Type de document</i> | <i>Extension</i> | | <i>Type de logiciel</i> |
|--------------------------------|---|---|--|
| texte ASCII | .TXT | | tout éditeur de texte |
| texte Word | .DOC | .RTF | Word ou un traitement de texte supportant le format RTF |
| hypertexte Web | .HTM | .HTML | Netscape lit ces fichiers |
| SGML | .SGML | | encore très rare, documents élaborés |
| Adobe Portable Document Format | .PDF | | Adobe Acrobat (installé) |
| Postscript | .PS | .EPS | lecteur Postscript type Ghost |
| image fixe | .GIF .BMP .TIFF ou .TIF .IMG | .JPEG ou .JPG .PCX .ART | Netscape sait lire ces fichiers Il faut un logiciel de dessin pour les modifier, type Photoshop, PaintshopPro |
| image animée | vidéo : animation : | .MPEG ou .MPG .VDO .AVI ou .FLI ou .MOVE | Quicktime VDO Live NetToob Stream |
| son | .AU .WAVE ou .WAV .AIFF ou .AIF .RA .RAM .XMD | .VOC .SND | Live Audio (installé) Real Audio Player (installé) Stream Works |
| images de molécules | .PDB .MDL | | Chemscape Chime (installé) Isis Draw (installé) |
| images 3D | .WRL .MOV | | Live 3D (installé) Quicktime VR |

Tableau 2 [12]

Il faut de plus en plus de logiciels utilitaires pour lire tous ces formats. Pour clarifier et faciliter la reconnaissance des types de fichiers, une norme a été mise en œuvre. C'est la norme MIME (Multipart Internet Mail Extension). Pour le multimédia, elle permet de faire correspondre les différents types de données aux applications assistantes ou external viewers.

2.2.1.2. Logiciels utiles pour la gestion du poste

2.2.1.2.1. Antivirus

Il existe 4 types de virus :

- les vers (worms)
- les bombes logiques (logic bombs)
- les bombes réglées dans le temps (time bombs)
- les chevaux de Troie (Trojan horses)

Ils peuvent infecter les secteurs de démarrage, les fichiers .COM et .EXE et les fichiers Word. Les virus peuvent être transmis par la messagerie électronique. Cependant les virus canulars (Hoax) encombrant les boîtes aux lettres. Les plus connus sont : Good Times, Greetings, Winzip300.

Pour prévenir tout problème, toujours vérifier les fichiers téléchargés AVANT l'installation, avec un logiciel antivirus :

- Mc Afee Viruscan
- Norton antivirus
- F-Prot
- Thunderbyte TbAv
- Microsoft Antivirus (livré en standard dans DOS 6.x)

2.2.1.2.2. Logiciel de décompression

Le plus courant : PKZIP, PKUNZIP, et sa version pour Win 95 : Winzip. Ce shareware décompresse les fichiers aux formats .ZIP .ARJ .GZ

2.2.1.2.3. Gestion des installations

Quand nous utilisons un ordinateur pour récupérer des informations via Internet encore plus que lors d'une utilisation "classiquement" bureautique, nous sommes amenés à installer un nombre toujours croissant d'applications assistantes afin de pouvoir exploiter tous types de fichiers. Il apparaît alors le problème des désinstallations des logiciels. Il est nécessaire d'utiliser un désinstalleur qui permettra de créer des archives de positionnement des fichiers des applications. Il faut réaliser cette opération sur un ordinateur ne possédant que le système d'exploitation. En effet, si le désinstalleur est mis en place après les applications, son efficacité se réduira à celle d'un effacement

manuel des répertoires donc en laissant un nombre conséquent de fichiers orphelins qui vont parasiter le disque dur.

L'application la plus connue parmi ces logiciels de nettoyage est : *CleanSweep*.

2.2.2. Recherche d'Information Scientifique et Technique

2.2.2.1. Ressources d'informations disponibles

2.2.2.1.1. Point d'entrée et carnet d'adresses

Le principe du Web est que chaque service d'information pointe sur les autres services du même genre ou traitant des mêmes sujets. Ainsi un service fait par une université pointerait sur ceux similaires d'autres établissements, ou bien encore un service consacré à la chimie par exemple signalerait souvent la liste des autres services intéressants dans le domaine. D'où la notion de point d'entrée : avec une adresse, nous pourrions déjà mener une recherche d'information parmi une multitude d'autres adresses, à priori intéressantes.

Les bons points de départ dépendent de notre sujet d'intérêt et de notre localisation : nous pouvons demander à nos collègues les bonnes adresses qu'ils connaissent mais surtout pêcher dans la littérature professionnelle que nous lisons régulièrement les adresses citées (c'est de plus en plus fréquent ; même les quotidiens et les magazines francophones le font). Pour bien utiliser un point d'entrée, il est préférable de bien connaître son contenu et d'avoir exploré ses différentes pages et rubriques.

Mais il faut se rappeler que l'objectif du point d'entrée (exemple du CHU de Rouen)⁹ est seulement de signaler les principales ressources dans un domaine précis. Il ne vise pas toujours à l'exhaustivité et n'est pas forcément mis à jour en temps réel car ses administrateurs ont souvent beaucoup d'autres choses à faire.

2.2.2.1.2. Répertoires géographiques

Les serveurs Web qui souhaitent être connus peuvent se déclarer dans leur pays d'origine auprès d'une institution qui en assure le recensement. En France, c'est l'Unité Réseau du CNRS (UREC)¹⁰ qui est chargé de ce travail. Ainsi l'UREC tient à jour la

⁹ CHU de Rouen - Ressources médicales : <http://www.chu-rouen.fr/ssm/watch.html>

¹⁰ UREC : <http://www.urec.fr/France/Web.html>

liste et la carte de France des serveurs. Les institutions homologues des pays connectés à l'Internet collaborent pour constituer un répertoire géographique mondial des serveurs Web. Il est ainsi possible de rechercher un service ou un serveur selon sa localisation géographique supposée.

Si nous nous souvenons par exemple que l'INIST est situé à Nancy en France, une recherche sur la carte de France des serveurs permettra de retrouver et de consulter son serveur Web qui donne beaucoup d'informations sur les produits et le fond documentaire de cet institut.[5]

La même recherche est faisable au niveau mondial.¹¹

2.2.2.1.3. Répertoires thématiques

Le fourmillement de services sur Internet a généré une nouvelle catégorie de services : les répertoires thématiques. Ils se distinguent des points d'entrée par leur couverture plus large mais non exhaustive. Leur objectif principal est plutôt de recenser les points d'entrée afin d'orienter les utilisateurs en fonction de leurs centres d'intérêt. Ainsi, des organismes ou des institutions coopèrent pour recenser par grands domaines scientifiques ou culturels les services les plus pertinents ou les plus utiles.

La Virtual Library¹² pilotée par le CERN à Genève - où est né le Web - est l'exemple type d'un projet "universel" de répertoire élaboré et mis à jour par des partenaires universitaires du monde entier.

Yahoo est un autre exemple de ces répertoires thématiques.

Ces répertoires sont constitués à partir d'explorations du réseau menées par leurs administrateurs, mais aussi grâce aux informations que leur envoient les producteurs de services. Certains présentent simplement une liste thématique des services, d'autres permettent en plus une recherche par mot clé. Les services qu'ils signalent sont en général des points d'entrée dans un domaine et rarement des services très spécialisés.

Jusqu'à récemment, les répertoires se concentraient sur les services de type W3. L'intégration continue des services Gopher, news et FTP dans le Web fait que les répertoires les signalent de plus en plus.

¹¹ UREC : http://www.urec.fr/France/Web_monde.html ; Virtual tourist : <http://www.vtourist.com/>

¹² Virtual Library du CERN : <http://www.w3.org/vl>

La Virtual Library du CERN par exemple, signale les principaux Newsgroups par domaine.

Lorsque nous souhaitons connaître les principales ressources disponibles en biologie ou en chimie, il faut consulter l'une de ces "Virtual Libraries". Nous y trouverons très rapidement un bon point d'entrée sur le sujet qui nous intéresse et que nous explorerons alors pour trouver les services plus spécialisés dans le domaine.

Fidèle à sa vocation de gardien de la flamme du Web, le CERN a réalisé une page qui recense toutes ces bibliothèques virtuelles¹².

2.2.2.1.4. Moteurs de recherche

Dans le cas d'une question dont le sujet est très pointu, l'approche par domaine est difficile. Dans ce cas, il faut interroger par mot clé les moteurs de recherche (search engines) pour identifier les pages ou documents du Web qui contiennent ce mot. Les moteurs de recherche sont des logiciels qui explorent le Web à intervalles réguliers pour identifier les documents proposés par les serveurs. Ils enregistrent l'adresse, les mots du titre, de l'entête et même pour certains, le contenu des documents trouvés ou seulement une fiche descriptive des documents. Ils constituent ensuite une base de données interrogeable par les utilisateurs du Web.

Il suffit de s'y connecter et de taper dans un formulaire de recherche quelques mots susceptibles de décrire ou de se trouver dans les documents recherchés. Leur interrogation fournit des listes d'adresses classées par ordre de pertinence selon des critères statistiques (fréquence d'apparition du terme cherché par rapport au nombre total de mots dans le document). La plupart des moteurs permettent d'utiliser les opérateurs booléens et les troncatures, mais selon des modalités différentes d'un moteur à l'autre. Ces moteurs sont encore en phase d'évolution pour la plupart. Ils ont chacun des règles de constitution et d'interrogation différentes (les résultats d'une requête sont donc différents de l'un à l'autre). Ils constituent cependant un outil fondamental pour explorer Internet, d'ailleurs tellement utilisé qu'à certaines heures de la journée (entre 10h00 et 12h00 ; et entre 16h00 et 18h00) il est très difficile de s'y connecter.

Les meilleurs indexent maintenant tous les types de ressources : Web, Gopher, les bases WAIS, les répertoires de fichiers récupérables par FTP ou les groupes de

discussion. Les plus connus ont pour nom Alta Vista, InfoSeek, Euroseek, Open Text, mais il en existe d'autres¹³.

Il existe un moteur qui ne recense que les ressources francophones : ECHO¹⁴.

La multiplication de ces moteurs a conduit à créer des méta-moteurs [9] permettant d'interroger en une seule étape plusieurs moteurs et d'éliminer les doublons parmi les réponses obtenues. Les réponses à une même question variant d'un moteur à l'autre [10], il faut souvent en essayer plusieurs, ce qui peut prendre beaucoup de temps. Les méta-moteurs n'apportent qu'une réponse partielle à ce problème car ils ne font qu'interroger en cascade plusieurs moteurs. De plus, tous les moteurs n'ayant pas les mêmes fonctionnalités, la question posée avec un méta-moteur ne peut pas utiliser simultanément, toutes les subtilités de syntaxe de tous les moteurs.

L'utilisation des méta-moteurs, n'est intéressante que pour cerner des serveurs généraux sur un sujet donné et pour donner les premiers éléments d'une recherche.[11]

2.2.2.1.5. Guides de ressources

Lorsque nous travaillons régulièrement sur un domaine scientifique il devient vite nécessaire de connaître non seulement les ressources Web mais aussi les principales listes de discussion, les newsgroups, les serveurs FTP et les serveurs très spécialisés accessibles sur abonnement. Dans ce cas, l'exploration systématique des répertoires et des moteurs de recherche est trop coûteuse en temps et il est plus judicieux de rechercher les guides de ressources sur le domaine.

Ces guides sont réalisés par des individus isolés ou par des équipes pour rendre service à la communauté et éviter à chacun de réinventer la roue. La plupart des guides sont d'abord publiés dans les listes de discussion ou dans les newsgroups spécialisés. Nous pouvons donc les retrouver dans les archives de ces groupes.

Cependant les guides se sont multipliés en même temps que les ressources d'information et pour permettre de les repérer et de les récupérer plus facilement il s'est créé un "guide des guides" qui constitue la référence incontournable : *The Clearinghouse¹⁵ for Subject-Oriented Internet Guides*, réalisé par l'université du Michigan et Argus Associates. Ce guide est organisé par thèmes ou interrogeable par

¹³ Liste exhaustive des moteurs de recherche sur le web : <http://www.beaucoup.com/engfren.html>

¹⁴ <http://www.echo.fr>

¹⁵ Clearinghouse Argus for subject oriented internet resource guides : <http://www.clearinghouse.net/>

mot clé. Il fournit la liste des meilleurs guides et permet même leur consultation en ligne ou leur récupération par FTP.

Selon le domaine étudié, les guides de ressources peuvent être très volumineux (plus de 50 pages de texte). Certains sont très sélectifs et ne signalent que les meilleurs services, d'autres sont exhaustifs, mais la plupart donnent un commentaire sur la qualité ou l'intérêt des services recensés.

2.2.2.1.6. Conférences électroniques

Avec les guides de ressources, nous voyons apparaître la valeur de réseau humain que constitue l'Internet : des spécialistes traitent et compilent des masses d'informations pour les rendre plus accessibles à d'autres spécialistes, intéressés par les mêmes sujets.

En utilisateur averti de l'Internet, nous devons nous intéresser aux modes de communication de groupe - les conférences électroniques - dont les formes principales sont les listes de discussion (mailing lists ou discussion lists ou encore listserv) et les groupes de discussion (usenet, newsgroups). Ces systèmes de conversation constituent à la fois le plus sérieux et le plus délirant des forums permettant de produire et de chercher de l'Information Scientifique et Technique.

Pour ne pas être submergé par le nombre et identifier rapidement les groupes pertinents, mieux vaut consulter les guides de ressources. Il existe une liste des conférences électroniques élaborée par l'équipe de Diane Kovacs¹⁶ qui est reconnue comme une référence internationale.

Une fois que nous avons identifié les groupes les plus intéressants dans votre domaine, il sera intéressant de les lire et de s'y abonner pendant quelques semaines pour juger de leur valeur. Rapidement la consultation des newsgroups et l'adhésion aux listes sont dévoreuses de temps. L'information n'y est pas toujours de qualité et la lecture du flot de messages quotidiens peut parfois dépasser les capacités d'un individu qui peut aussi avoir d'autres tâches. Il sera alors temps de se limiter aux groupes dits modérés, dans lesquels l'information est filtrée - mais non censurée - par un modérateur, améliorant la pertinence et réduisant le flux.

¹⁶ D.Kovacs et al., Directory of Scholarly Electronic Conferences. 10th Ed, 1995 :
<http://www.n2h2.com:80/KOVACS/>

Pour toute personne ayant à mener des recherches d'information sur des sujets variés, un bon compromis consiste à suivre les groupes spécialisés dans les sources d'information (par exemple CHMINF qui se concentre sur les sources d'information en chimie) et les groupes rassemblant les documentalistes et les professionnels de l'information (biblio-fr, adbs-info en France ou PACS-L, Web4lib dans le monde, pour ne citer que les plus connus). En participant à ces groupes on reçoit continuellement une information sur les meilleures ou les nouvelles ressources du domaine. Dans le cas où l'on doit répondre à une question sur un sujet mal connu, on peut faire appel à ces groupes pour demander si quelqu'un connaît la réponse ou peut donner des pistes de recherche, des références et même fournir des documents.

Avant de poser une question à un groupe, ou si l'on a un souci de confidentialité, il faut toujours vérifier dans les archives des listes et dans les FAQ¹⁷ des newsgroups que le sujet n'a pas déjà été traité. En effet, dans ce domaine pourtant très informel, la communauté Internet s'est organisée pour faciliter l'accès à sa " mémoire ". Il est donc possible d'interroger par mots clés sur différents serveurs les archives et les FAQ de pratiquement toutes les conférences électroniques.

Pour accéder aux FAQ, nous pouvons nous connecter sur presque tous les serveurs Web universitaires qui les proposent en consultation ou sur un service spécialisé d'archives de FAQ. La liste des listes francophones¹⁸ est présentée et mise à jour à l'Université de Rennes. Le service proposé permet de s'abonner, et pour certaines listes de discussion, de consulter les archives et de connaître la liste des abonnés.

Inconvénient : Tout en générant des pratiques collaboratives, la communication de groupe provoque un éparpillement de l'information (nombre de groupes élevé) et empêche sa validation, ce qui est particulièrement gênant pour l'information scientifique et technique. Les cas sont déjà nombreux d'informations fausses ou de mauvaise foi diffusées dans des conférences électroniques. A cela s'ajoute la redondance des informations trouvées : les mêmes données se trouvent à plusieurs endroits, parfois tronquées, parfois sans mention d'auteur et dans des versions obsolètes.

Il faut donc être extrêmement attentif à la validité des informations recueillies si elles doivent aider à la prise de décision (orientation de la politique de recherche d'un

¹⁷ voir note n°1 page 13

¹⁸ Listes de discussion francophones, FRANCPHOLISTES : <http://www.cru.fr/listes/>

laboratoire, choix d'équipement, recherche de partenaires). Une règle de base est le recoupement de l'information avec d'autres sources que l'Internet.

2.2.2.1.7. Signalement des nouveautés

L'Internet est vivant : chaque jour des services naissent, déménagent, meurent. Pour permettre aux utilisateurs de suivre plus facilement l'utilisation du réseau, divers organismes se sont donné pour tâche de signaler les nouveaux services qui apparaissent sur le Web. Le CNRS réalise ce travail pour toutes les nouveautés françaises¹⁹. Au niveau mondial, dans le domaine scientifique, il faut aussi signaler l'Internet Resource Newsletter²⁰. Ces deux services reçoivent directement l'information des producteurs et explorent le réseau pour identifier les nouveautés.

2.2.2.2. Méthode générale

Compte tenu de la structuration actuelle des services, il est déjà possible d'élaborer une sorte de méthode générale de recherche d'IST sur l'Internet. L'objectif de ce paragraphe est donc de présenter les grandes étapes d'une recherche d'information, permettant de gagner du temps, mais sans garantir que nous trouverons effectivement une réponse à toutes les questions. En s'inspirant de cette démarche, nous pourrons plus facilement acquérir les connaissances et les réflexes de base qui permettront de devenir un Internaute expérimenté, pour envisager alors une méthode de recherche plus fine et mieux adaptées à des questions précises.

Avant de nous lancer dans une recherche, il nous faut clairement poser la question (sujet, champ d'application, type de données cherchées)²¹ afin de décider si c'est bien une recherche dans Internet qui vous fournira la réponse attendue. Ce préliminaire peut sembler superflu, mais l'engouement actuel pour le réseau fait que l'on perd son sens critique et que l'on en oublie les bases de données bibliographiques ou même les fichiers de la bibliothèque.

Si c'est le cas, il va nous falloir trouver un point d'entrée qui signalera les services spécialisés sur le sujet qui nous intéresse. Pour ce faire nous irons consulter les répertoires thématiques (comme la Virtual Library du CERN ou Yahoo ou le Web du

¹⁹ Nouveaux serveurs en France : <http://www.urec.fr/France/dnf.html>

²⁰ Internet Ressource Newsletter : <http://www.hw.ac.uk/libWWW/irn/irn.html>

²¹ John S.Makulovic. "10 Tips on managing your Internet searching", Online, 1995, 19(4) - pp 32-37

CNRS pour la France). Nous obtiendrons en principe les meilleurs points d'entrée dans le domaine. A nous de les explorer pour trouver ceux qui semblent le mieux répondre, en n'oubliant pas de mémoriser leur adresse. Rappelons que c'est par la connaissance approfondie du contenu d'un point d'entrée que nous saurons atteindre le service qui répond à notre question.

Si cela ne suffit toujours pas pour trouver l'information cherchée ou si notre question est très pointue, nous irons interroger les moteurs de recherche par le terme le plus spécifique de notre question. Ils fourniront certainement des points d'entrée et parfois même les meilleurs adresses sur un sujet, mais leur utilisation ne garantit pas non plus l'exhaustivité. A nous de bien connaître le mode d'emploi de deux ou trois moteurs et d'évaluer la qualité des réponses qu'ils renvoient.

Si nous avons à répondre à des questions sur un domaine scientifique, nous irons chercher un guide des ressources sur le sujet à l'université du Michigan dans la liste des guides appelée Clearinghouse²². Nous passerons quelque temps à tester les adresses signalées dans ce guide et finirons par avoir une bonne connaissance des ressources dans le domaine.

Enfin, si nous devons assurer une veille en continu, la visite régulière des sites sélectionnés lors de recherches antérieures, est indispensable. De plus, il est nécessaire d'utiliser les services d'un "agent intelligent". Ce type de logiciel permet de sélectionner les informations que nous souhaitons retrouver, en remplissant un formulaire. Ensuite, le logiciel se charge de télécharger tous les jours les nouvelles concernant nos domaines d'intérêt.

Utilisant beaucoup la messagerie électronique, nous travaillerons certainement autant avec d'autres experts humains qu'avec des services d'information. Nous consulterons les services de signalement des nouveaux services pour nous tenir au courant, notre carnet d'adresses nous permettra de répondre à la plupart des questions.

²² <http://www.clearinghouse.net>

2.2.2.3. Outils d'aide à la recherche

2.2.2.3.1. Agents veilleurs [14]

La dernière version du navigateur Netscape se nomme Communicator 4.01 et intègre un nouvel outil : Netcaster. Il s'agit d'un agent "veilleur". Il est possible de sélectionner les différents canaux d'informations et de créer son propre profil d'interrogation.

Ces dernières versions des navigateurs sont des bêta-versions non encore complètement fiables. En effet, toutes les erreurs ou bogues de fonctionnement ne sont jamais tous éliminés lors de la sortie du logiciel. Il faut attendre des mises à jour et finalement la version stabilisée du produit pour bénéficier d'une utilisation efficace et sans problème.

Afin de se prémunir des "erreurs de jeunesse" de ces nouvelles possibilités des navigateurs, nous pouvons lancer l'exécution simultanée d'un agent "veilleur" et du navigateur. L'offre de tels outils est de plus en plus importante, tant au niveau du nombre de ces agents sur le marché que de leurs fonctions. Nous avons utilisé PointCast Network²³ et Backweb²⁴ en association avec Personal InfoSeek²⁵. Cette catégorie d'agents a été créée afin de satisfaire les besoins en veille technologique et en intelligence économique. Grâce aux agents "veilleurs", nous pouvons réaliser des profils d'interrogation contenant nos critères d'intérêt. Cela permet de se tenir informer sur les domaines qui nous concernent vraiment, sans devoir parcourir systématiquement plusieurs sites Web fournisseurs de news²⁶.

2.2.2.3.2. Agents aspirateurs ou Off-line browsers

Ce type d'outil permet de connaître toute l'arborescence et la structure des répertoires et fichiers d'un site. De plus, il est possible d'obtenir une copie exacte du site sur le disque dur de notre propre machine, donc de consulter le site en n'étant pas connecté (1^{er} intérêt) et à une vitesse bien supérieure à celle rencontrée sur le réseau avec une connexion par modem (2^{ème} intérêt). En effet, le passage d'un lien à un autre est instantané!

²³ PointCast Network : <http://www.pointcast.com>

²⁴ BackWeb : <http://www.backweb.com/home.html>

²⁵ Personal InfoSeek : <http://personal.infoseek.com> : permet de recevoir les informations par e-mail.

²⁶ Wire services : <http://www.vgol.com/wire.html>

Ces "aspirateurs" de site peuvent rendre de grands services en permettant un gain de temps important sur les sites les plus fréquemment utilisés, à condition toutefois de connaître leurs limites.

Il existe deux type d'organisation pour un site Web :

- Cas n°1 : tous les liens apparaissent clairement sur les pages du site. Dans ce cas, l'agent "aspirateur" est opérationnel à 100%. En effet, le site sera recréé à l'identique sur le disque local. Tous les liens sont alors utilisables comme lors d'une session en ligne, mais leur activation est immédiate.
- Cas n°2 : le site est pourvu d'un moteur de recherche qui permet d'interroger la base de données interne, locale. A ce moment, tous les liens du site ne sont pas affichés car l'arborescence des fichiers n'est pas hiérarchique et séquentielle. Dans ce cas, l'agent "aspirateur" ne peut pas rapatrier en local toute la base de données qui fonctionne en amont du site.

L'analyse du site par le moteur peut se faire sur une journée complète! Il faut donc lui préciser jusqu'à quelle profondeur il devra "aspirer" les liens. Il est utile aussi de choisir l'adresse d'entrée dans le site. D'autres options sont offertes afin de réaliser un travail plus fin de sélection sur les parties du site que nous désirons dupliquer. En effet, à moins de disposer de moyens de stockage très importants, il est nécessaire de surveiller le remplissage de notre disque dur local. Dans ce but, il est possible de préciser au logiciel les types de fichiers qu'il doit enregistrer ; à nous de les avoir identifiés au préalable.

De tels outils peuvent également être employés pour assurer une veille technologique : il suffit alors de se connecter aux sites appropriés et d'y récupérer les informations nouvellement mises à jour [13]. Dans ce contexte, l'utilisation des outils off-line s'apparente aux nouvelles fonctions de *push* proposées dans les prochaines versions de Netscape ou d'Internet Explorer, à cette différence près que c'est l'utilisateur lui-même qui décide des périodicités de mise à jour (le *push* décide arbitrairement de la mise à jour sans que l'utilisateur puisse intervenir).

2.2.2.3.3. Agents de recherche intelligents

Il en existe deux types :

- Automate : après avoir choisi les moteurs de recherche "classiques" qui constitueront les ouvriers de la recherche, nous introduisons les mots clés de la question. L'automate enregistre les résultats sur le disque dur et élimine les doublons afin de nous faire gagner du temps.
- imitateur : le principe est le même que pour les "automates", mais ils constituent une deuxième génération. Ils se greffent sur le navigateur et observent toutes les opérations que nous réalisons sur Internet. Exemple : *More Like This* permet, une fois que nous avons trouvé quelque chose d'intéressant, de retrouver des documents similaires. Son fonctionnement est simple et il est très utile dans les recherches quotidiennes. Néanmoins, ce logiciel génère beaucoup de bruit par réponse.

2.2.2.3.4. Exemple d'aide à la recherche interne : *Excite for Web Servers*

Les sites devenant de véritables labyrinthes, et les internautes ne supportant plus de perdre du temps à errer avant de retrouver l'information, de plus en plus de sites offrent une option de recherche sur leurs pages web grâce à un moteur de recherche interne. Très souvent, cet outil est produit par Excite. *Excite for Web Servers* fournit aux utilisateurs un accès à un moteur de recherche basé sur les concepts et non sur les mots clés. Il permet donc de retrouver des documents contenant l'idée de la question mais ne comportant pas forcément les mots clés spécifiés. Même si la base du moteur est complexe, les utilisateurs peuvent poser leurs questions simplement et quasiment en langage naturel. Plus la question comportera de détails, plus la réponse sera proche du résultat optimal. Il ne faut pas hésiter à utiliser de nombreux mots d'interrogation pour décrire notre besoin. Comme dans tous les outils de recherche, les opérateurs booléens sont disponibles afin de poser des questions plus précises.

Chapitre III : L'INTERNET POUR LE C.R.P.F.

1. L'INTERNET POUR LES "UTILISATEURS

NON-DOCUMENTALISTES"

La maîtrise des différents outils utilisés lors d'une recherche peut prendre quelques mois de pratique. Nous pouvons considérer qu'une personne qui arrive à ce niveau de compétence est quelqu'un de rare. En effet, plusieurs enquêtes et témoignages [7] montrent que la plupart des utilisateurs de l'Internet, le plus souvent des chercheurs, connaissent quelques ressources dans leur domaine, souvent glanées au hasard ou transmises par des collègues plus avancés, mais qu'ils n'ont ni le temps ni le besoin d'explorer davantage les ressources du réseau. Ils consultent toujours les mêmes services d'information, en très petit nombre, qui répondent à leurs besoins quotidiens et utilisent l'Internet plutôt comme moyen de communication, grâce à la messagerie électronique. La plupart des utilisateurs, restent assez démunis devant des questions qui sortent de leur champ d'étude. Nous voyons donc que le rôle du documentaliste ou de spécialiste de l'information reste essentiel pour faire le tri dans la masse des informations disponibles, pour signaler et présenter les ressources les plus importantes dans un domaine, et mener des recherches à la demande de ces publics.

Le centre de documentation pourra mettre en place un serveur Web constituant un point d'entrée sur les différents domaines de compétence. Il y a là un impératif économique (une personne consacre du temps à cela et en fait bénéficier les autres) et professionnel (d'autres ressources sont explorées et le champ de vision est élargie).

2. POLITIQUE GLOBALE CHEZ PIERRE FABRE

2.1. L'Internet aujourd'hui

2.1.1. Courrier électronique au CRPF

2.1.1.1. Etude de l'existant

L'Internet au CRPF :

Connexions effectives :

- Direction du CRPF : possède ses propres connexions à l'Internet
- Centre de Documentation Scientifique
- Biométrie-Informatique de recherche

Ces deux dernières unités ont été équipées de façon autonome par nécessité de fonctionnement.

De nombreuses connexions restent à réaliser :

Au total, 240 personnes sont concernées par la mise en place de ce nouvel outil.

2.1.1.2. Besoins

La messagerie est le service le plus utilisé d'Internet et qui justifie à lui seul l'abonnement de nombreux utilisateurs au réseau.

⇒ Il s'agit d'un moyen simple et très économique de communication entre personnes. La messagerie permet d'adresser pour le coût d'une communication locale, un message à toute personne ayant une boîte au lettre électronique, quelque soit son lieu physique.

⇒ L'e-mail n'est pas limité à l'envoi des textes : il peut ainsi transmettre en fichier annexé tout document pouvant être converti sous forme numérique (photo, vidéo, son).

⇒ Les logiciels de messagerie électronique permettent de façon simple de relever sa boîte au lettre (fréquence personnelle), de répondre et d'envoyer des messages dans le monde entier. Envoyer le même message à une ou plusieurs personnes demande le même temps.

Déjà, par l'intermédiaire d'une structure temporaire, le Centre de Documentation Scientifique gère des demandes concrètes comme des courriers réguliers, des demandes d'enregistrement d'articles.

2.1.1.3. Solutions à envisager

2.1.1.3.1. Adresse unique

Deux solutions sont envisageables pour gérer une seule adresse pour l'ensemble du centre de recherche :

1. Un poste centralisé avec une personne qui se charge d'envoyer les messages du CRPF et de recevoir donc de trier les messages reçus. Cela implique l'impression et la réexpédition vers les services.

Avantage : Manipulation quotidienne de l'outil donc une bonne connaissance des procédures ; résolution rapide des problèmes et gain de temps.

Inconvénient : Beaucoup de temps nécessaire à la gestion de ce service par une seule personne.

2. Un poste en accès libre

Avantage : Gain de temps pour le service documentation car c'est l'utilisateur final qui réalise les opérations de gestion de ses propres messages.

Inconvénient : Engorgement du service en libre accès si plusieurs utilisateurs arrivent en même temps (penser à téléphoner avant de se déplacer pour s'assurer que le poste est accessible).

Possibilité d'effacement et de perte de message par réalisation de fausses manoeuvres ou d'erreurs sachant que l'utilisation sera ponctuelle contrairement à la première solution.

Trois autres points sont à mettre au passif d'un système fournissant une seule adresse pour un nombre élevé de personnes :

- Difficultés pour identifier les messages concernant le département ou la division liée à un nombre important de messages affichés les uns à la suite des autres.

2.1.1.3.2. Adresse partagée

Une mise en oeuvre à court terme, peut être envisagée par recours à un prestataire extérieur : à partir d'une seule ligne téléphonique Numéris, nous pouvons avoir jusqu'à 20 connexions simultanées (Web et e-mail), et créer nos propres adresses. Ce système est protégé par un firewall et est compatible Windows 3.xx, 95 et NT 4.0. Cette prestation est bien sûr payante.

2.1.2. Recherche sur le Web

Pendant la durée du stage, un accès aux ressources du Web a été proposé à tous les chercheurs du CRPF sur un ordinateur en libre service au Centre de Documentation Scientifique.

2.2. Solution future

Si la mise en place de la solution e-mail "à adresse unique" peut se faire rapidement en tenant compte du temps de formation à l'outil, l'évolution passera obligatoirement par une multiplication d'adresses électroniques pour toutes les personnes intéressées dans chaque division. Cette solution doit passer par la procédure globale du Groupe.

La Direction de l'Organisation des Systèmes d'Information est en charge de la politique Internet du Groupe. La solution préconisée passe par des connexions via le réseau interne, nécessitant des spécifications particulières du poste client. Cette solution est effective, fonctionnant déjà sur certains sites, mais ne comprend pas le courrier électronique. Le déploiement du courrier électronique ouvert sur l'extérieur semble n'être envisagé qu'après la mise en place de la messagerie interne (Lotus Notes-Domino), soit après 1998. De plus l'ouverture en nombre à l'Internet pose encore quelques problèmes de capacité volumétrique.

3. PRINCIPALES SOURCE EN CHIMIE ET BIOLOGIE SUR

L'INTERNET

| | |
|--|--|
| <i>Chimie & Biochimie (niveau universitaire)</i> | |
| Voir Chap II, p15 | |
| La plupart des cours de biologie sont recensés à : BioChemNet, a directory of biology and chemistry educational ressources : http://schmidel.com/teaching.htm | |

| | |
|--|---|
| <i>Les ressources en chimie</i> | |
| Documents de l'IUPAC : nomenclature des composés et enzymes, glossaires de termes, poids moléculaires. | http://alpha.qmw.ac.uk/~ugca000/iupac.html/ |
| Thesaurus des termes chimiques (NASA) | http://www.sti.nasa.gov:80/thesaurus/C/word2498.html |

| | |
|---|--|
| <i>Systèmes de publications en ligne</i> | |
| Voir Chap II, p15 | |
| PubMed (NLM) fait un lien entre la base Medline et le texte électronique des publications : http://www4.ncbi.nlm.nih.gov/PubMed/ | |

| | |
|---|---|
| <i>Journaux en ligne et journaux électronique</i> | |
| Partie Internet de Trends in Analytical Chemistry | http://www.elsevier.nl:80/inca/homepage/saa/trac/frames.shtml |
| Une liste des journaux en chimie chez BetaCyte Consultants. Signale près de 300 titres électroniques classés selon leur coût. <u>Très bien fait.</u> | http://www.betacyte.com/ |
| Journal uniquement disponible sous forme électronique : Neuroscience-net | http://www.neuroscience.com/ |
| Experimental Biology Online - EBO : Springer édition (gratuit pour 2 ans) | http://science.springer.de/ebo/ebo-main.htm |

| | |
|--|---|
| <i>Les grands éditeurs en chimie proposent des accès intéressants aux sommaires et à des "supplementary materials"</i> | |
| American Chemical Society | http://pubs.acs.org/journals/jrnhome.html |
| Royal Society of Chemistry | http://chemistry.rsc.org/rsc/j1.htm |
| Society of Chemical Industry | http://ci.mond.org/ |

| | |
|--|---|
| Conference Listing (partie chimie) | http://hackberry.chem.niu.edu:70/0/ConferenceListings/SpecialistMeetings/index.html |
| Toutes les conférences scientifiques référéncées à Berlin | http://www.chemie.fu-berlin.de/chemistry/index/conf/ |
| ACS Calendar of Meetings | http://www.acs.org/meetings/calendar/calendr1.html |
| Chemical & Engineering News : Calendar of Meetings & Events | http://pubs.acs.org/journals/scheds/events.html |
| Liste en biologie réalisée par Infobiogen | INFOBIOGEN (fr) - Rubrique congrès : http://www.infobiogen.fr/gen/them3.html#CON |

Bases de données

Les bases de données commencent à se multiplier. En chimie, la plupart sont en accès **payant ou sur abonnement**.

Joseph Warden propose un lien vers une page de présentation des principales bases accessibles :

http://www.rpi.edu/dept/chem/cheminfo/chemres/chemres_03.html

Bases bibliographiques

| | |
|--|---|
| Les grands éditeurs proposent des <u>accès payants</u> , mais avec des interfaces simplifiées exemple type :  | http://stneasy.cas.org/ |
| Un exemple de base bibliographique de 17.000 sommaires de périodiques avec commande intégrée des documents : UNCOVER | http://www.carl.org/uncover/unchome.html |
| MEDLINE est accessible gratuitement sur 17 sites. Le service qui offre les plus puissantes possibilités de recherches est : Evaluated Medline | http://biomednet.com/db/medline |

| <i>Bases numériques et factuelles</i> | |
|---|---|
| Web Elements propose les données fondamentales sur les éléments chimiques | http://www.shef.ac.uk/~chem/web-elements/ |
| Une base en accès libre après déclaration : Base de données sur les phases stationnaires chirales : une petite partie de CHIRBASE : | MOLECULAR DATABASE FOR CHIRAL CHROMATOGRAPHY (à Marseille) : http://u3pic133.u-3mrs.fr:8080/chirbase/welcome.htm |
| Le serveur Titane héberge: Le Beilstein, la Cambridge Structural Database, et d'autres banques (payant). | SERVEUR NATIONAL DES BASES DE DONNEES EN CHIMIE DU CNRS http://vulcain.u-3mrs.fr:8000/stt |
| Fundamental Physical Constants | http://physics.nist.gov/PhysRefData/codata86/codata86.html |
| Serveur et liens sur la RMN | Centre de Biochimie Structurale de Montpellier (CNRS - INSERM) http://tome.cbs.univ-montp1.fr/RMN/base.html |
| Un catalogue des bases de données de spectres : listes de bases payantes | Chemical Reference Spectra http://www.intaccess.com/fdm |
| Base des formules 3D | La MathMol library contient les structures 3D de nombreuses molécules biologiques et chimiques. Cette liste vient juste d'être mise en place. Ce site est surtout orienté vers l'enseignement. http://www.nyu.edu/pages/mathmol/library/library.html |
| Le Vidal du particulier | http://www.vidal.fr/ |
| BIAM : cette banque fournit des informations sur les interactions entre médicaments. | http://www.cri.ensmp.fr/biam/acceuil.html |

| <i>Brevets et Recherches en cours</i> | |
|--|---|
| brevets en texte intégral : payants : | Questel QPAT US : http://www.qpat.com/ Derwent, Patent Explorer : http://www.dash.uk.inter.net |
| gratuits : | IBM : http://patent.womplex.ibm.com/ |
| Recherches en cours à l'INSERM | http://inserm.cnusc.fr:8113/ |

Groupes de biologie (Biosciences)

Voir Chap II, p16

Les groupes de biologie sont regroupés sous BIOSCI / BIONET : <http://www.bio.net/>

Structures d'entreprises chimiques, réseaux d'entreprises

Nous pouvons interroger
par nom de société ou de
produit

Elles sont recensées à l'adresse : <http://www.chem.com/>

Grands organismes, agences gouvernementales et laboratoires de recherche

Pour le monde : Chemdex

<http://www.shef.ac.uk/~chem/chemdex/>

pour la France : UREC

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

Serveurs de Biologie,
Génétique &
BioInformatique en France
et Francophonie

Liste du CHU de Rouen :
<http://www.chu-rouen.fr/ssf/biofr.html>

Sécurité au laboratoire - Données toxicologiques

Le meilleur site au monde
pour les questions de
sécurité

SIRI - Vermont ressources on Safety : <http://hazard.com/>

Logiciels de chimie et biologie

En France, le centre
documentaire
d'information et
d'éducation en Chimie à
Nice propose un catalogue
très important des logiciels
disponibles

<http://www.unice.fr/cdiec/>

index des logiciels
présenté par Chemdex

<http://www.shef.ac.uk/~chem/chemdex/software.html>

Logiciels destinés à la
biologie moléculaire

Une liste exhaustive proposée par l'Institut Pasteur :
<http://www.pasteur.fr/other/biology/francais/bio-software-fr.html#Packages>

Banque de logiciels du
CDIEC

<http://www.unice.fr/cdiec/>

| | |
|---|--|
| <p>Dédié à la chimie française, ce site a pour objectif d'aider tous les acteurs, du professeur d'université au recruteur, de l'élève au PDG, à trouver les informations qu'ils recherchent.</p> | <p>Chimie France : http://www.mygale.org/05/nic55/</p> |
|---|--|

Tableau 3 [21]

Chapitre IV : L'INTERNET EST UNE VITRINE

L'élaboration d'un site groupe n'est évidemment pas du ressort du Centre de Documentation. En revanche, profiter d'un Intranet pour créer un site Documentation semble une éventualité à envisager. Il existe aujourd'hui des interfaces entre serveurs HTTP et les structures de données comme les SGBD qui permettent de construire des services Web dont la mise à jour est dynamique. Ces interfaces offrent ainsi aux centres de documentation et aux services de diffusion de l'information, la possibilité de mettre en accès sur le Web leurs banques de données.[15]

La création d'un serveur Web peut devenir pour un centre de documentation scientifique un enjeu important. Ce média est simple d'emploi, convivial et multimédia. Le Web devient pour les individus qui savent bien l'utiliser une source d'information irremplaçable. Mais, avant toute réalisation d'un tel service d'information, il faut engager une réflexion sur les principes juridiques qui régissent l'Internet, sur les moyens sécuritaires qu'il convient de mettre en place, sur la structure et l'ergonomie du service que nous voulons implanter et sur les outils que nous voulons utiliser.

1. DROIT ET SECURITE

L'Internet est peut-être indépendant, mais il n'est pas sans loi, et de nombreuses réglementations lui sont applicables. Avec le droit français, les conventions internationales ou les directives communautaires, il est possible de s'appuyer sur une mosaïque de textes qui réglementent la circulation et les échanges de données. Ainsi, le prétendu "vide juridique" concernant l'Internet n'est pas aussi profond qu'on le croit souvent.[6]

La sécurité est un point critique pour tout réseau. En effet, l'ouverture massive que supposent les autoroutes de l'information renforce les menaces à l'encontre de la confidentialité, de l'intégrité et de la disponibilité des données, notamment par des intrusions distantes. La sécurisation des transmissions contre les indiscretions et les détournements ainsi que la protection des accès sont des points capitaux qu'il convient de prendre en considération.

Il faut savoir qu'il existe de nombreuses solutions techniques pour assurer l'authentification des correspondants et la confidentialité des informations transmises

par cryptage. Le cryptage ou chiffrement transforme les données en une forme lisible seulement à l'aide de clés [18]. En outre, l'administrateur-système, peut utiliser un "firewall" pour filtrer les données qui transitent entre l'Internet et le réseau local dont il est responsable. Un firewall est au choix : un chien de garde, un garde-barrière, un mur pare-feu, ou tout procédé qui permet de contrôler les transactions entre un réseau d'entreprise et l'extérieur. Il permet de vérifier que les informations confidentielles ne sont pas consultées, que des transactions douteuses ne sont pas opérées, ou que ni l'ensemble du réseau ni l'un des serveur n'est mis à mal [19]. Enfin, même si les firewall sont des éléments clés de la sécurité d'un réseau privé raccordé à l'Internet, ils ne doivent pas être la seule ligne de défense, car il ne faut pas oublier que la plupart des problèmes de piratage proviennent d'actes malveillants commis en interne.

2. CONCEPTION

Une réflexion sur le fond ainsi que sur la forme du service Web doit ensuite être menée. Elle peut être supportée par la visite et l'analyse de site déjà existants ayant les mêmes activités dans le domaine de la documentation scientifique (INRIA, INIST, British Library, BNF).

Il est important de garder à l'esprit que l'Internet est un support qui suit des règles de fonctionnement particulières et que la création d'un bon site Web ne s'improvise pas [20] :

- ⇒ la définition du contenu de ce service, des éléments d'information que l'institution veut mettre en ligne est bien sûr essentielle ; de même que la détermination des cibles et des enjeux.
- ⇒ vient ensuite une phase de préparation à la réalisation, permettant de déterminer le parti que l'institution peut tirer d'un site Web, et pour quel coût.
- ⇒ puis une phase de réflexion sur la conception du site : définition de l'arborescence du site, définition d'une charte graphique, réflexion sur l'ergonomie, choix des standards.
- ⇒ la réalisation proprement dite du site consiste à intégrer les divers éléments, préalablement sélectionnés en utilisant le langage HTML.
- ⇒ après avoir choisi le mode d'hébergement et mis le service en accès sur le réseau, il faut enfin se préoccuper de la promotion du site.

3. EVOLUTION TECHNIQUE

Le phénomène d'intégration des technologies existantes par celles déjà présentes sur le Web n'en est probablement qu'à son commencement. D'ici peu, tous les outils de développement client-serveur, tous les SGBD, tous les systèmes de gestion de bases documentaires, toutes les applications bureautiques seront ouverts sur le Web, et cela sans aucune restriction.[16]

CONCLUSION

En simplifiant les langages d'interrogation des bases de référence on-line, l'Internet doit trouver sa place sur le bureau du *end-user* ou utilisateur final, afin de répondre aux questions ponctuelles courantes qu'il est amenées à se poser au cours de ses recherches.

Le rôle du documentaliste professionnel réside alors dans la réponse à des questions nécessitant une stratégie de recherche poussée en utilisant des moyens plus performants que l'Internet en ce qui concerne l'exhaustivité et la validité des informations (recherche d'antériorité brevet par exemple). Aujourd'hui, le documentaliste doit être à même de répondre aux questions des utilisateurs de l'Internet. Il doit posséder une longueur d'avance et passer du temps à tester les nouveaux outils qui ne cessent d'apparaître à propos de l'Internet. En effet, l'utilisateur final est consommateur alors que le documentaliste est un fournisseur de services clé en main.

L'Internet sera donc intégré dans le Centre de Documentation Scientifique sans entraîner la désintégration de celui-ci.

ANNEXE : SIGNETS

crpf3's Bookmarks

Actualités sur le Web

Chronique de Cybérie

Actualité du Web sur le Web

MOTEURS DE RECHERCHE

Anglophones

Alta-Vista

Robot ; Très grande richesse des possibilités de recherche formelle.

The Argus Clearinghouse

Propose des guides thématiques, offre une interrogation par mots clés.

Infoseek

Index compilé par un robot.

Infoseek Guide

Recherche dans le web, les newsgroups et les pages ftp.

Possibilité d'adhérer à Infoseek Professional qui exécute des recherches sur les sites du business.

The Open Text Index

Donne de bons résultats. La syntaxe est assez succincte mais le masque de saisie des questions est pratique et efficace.

Hotbot

Robot

Excite

Index compilé par un robot.

version française

Webcrawler

Index compilé par un robot

Magellan

Index administré manuellement.

A chaque site retrouvé est associé un commentaire de l'équipe de Magellan. Cette option crée une valeur ajoutée pour ce moteur car elle permet d'éliminer des réponses rapidement. Il s'agit d'une évaluation éditoriale précise de chacun des sites référencés.

Adapté pour localiser le site d'une entreprise, pour des renseignements sur un secteur d'activité utiliser Yahoo.

Lycos

Index compilé par robot

Yahoo

Index administré manuellement et qui joue la carte de l'exhaustivité. Intéressant pour obtenir des infos non pas sur une société précise (Magellan), mais sur un secteur d'activité.

Européens

Euroseek

Offre la possibilité de pouvoir sélectionner le pays et la langue des documents.

Francophones

Echo

Robot, permet de rechercher les sites pointant vers une adresse française en cochant l'option "des liens vers".

Il indique le nombre de liens, comme le moteur français Nomade, pour identifier les sites fédérateurs.

les sites fédérateurs.

Explore également les News, e-mail francophones.

Ce site est édité par la société France Télécom. On y retrouve les "pages Web" ou pages Zoom : actuellement, pages jaunes de l'annuaire de France Télécom.

Bientôt disponible : les pages blanches. Les deux autres services ne sont pas très exhaustifs, les pages marques offrent cependant une liste alphabétique de marques exerçant une activité en France.

Nomade

Index administré manuellement.

Outil de type annuaire.

Présentation similaire à celle de Yahoo : classement par thèmes.

Indication de la pertinence et du nombre de liens de chaque site.

Trouvtout

Index administré manuellement.

Listes peu exhaustives, mais permet la visualisation alphabétique des sites français recensés.

Image et Son

Web Seer

Version française. Permet de préciser la recherche en indiquant le type d'image.

WebSeek

Recherche thématique.

Lycos son et image

Yahoo image surfer

Recherche thématique et par mots clés.

Ressources images sur Internet

un répertoire des adresses à connaître pour trouver des images sur tous les thèmes

Off-line Browser

Aspirateur de site Web

Teleport Pro

Shareware

liste de moteurs

Informations économiques

Advertising Age

Business Yahoo

Rubrique économique de Yahoo

Bizweb

Liste hiérarchique de thèmes et possibilité d'interrogation par mots clés

Companies on the Net

Annuaire de sites pouvant être interrogés dans le domaine économique : EDGAR, Thomas Register, Bizweb...

Infos économiques extérieures

Direction des Relations Economiques Extérieures

Forum & News

Déjà News

Logiciel de consultation de News.

InfoScan

Logiciel de filtrage et de visualisation de l'information issue du courrier électronique et des News.

Outils de Filtrage

BackWeb

éditeur de nouvelles quotidiennes

éditeur de nouvelles quotidiennes

Peut être utilisé avec le profil Personal Infoseek, ce qui permet de rapatrier les news sur son écran de veille. De plus les infos sont collectées lors des temps morts de connexion, cela permet de ne pas ralentir les connexions principales.

Personal Infoseek

réponse personnalisée envoyée soit dans la boîte aux lettres, soit sur une page web "personnelle" réalisée par BackWeb.

Pointcast network

L'un des premiers systèmes gratuits à offrir une information personnalisée.

My Yahoo

News personnalisées

stanford netnews filtering service

édition de profils à partir d'informations extraites des groupes de News.

Méta-moteurs

Outils de recherche qui intègrent l'interrogation de plusieurs moteurs.

Attention, la typographie de la question sera toujours plus simple que lors de l'utilisation d'un seul moteur. En effet, tous les moteurs de recherche ne fonctionnent pas avec le même "langage" d'interrogation donc toutes les options ne sont pas disponibles sur un méta-moteur. L'intérêt se trouve lors d'une question vague : ces outils permettent de cerner grossièrement les sites et les mots clés intéressants. Ensuite, il est nécessaire d'utiliser un moteur plus exhaustif dont la syntaxe est maîtrisée.

Metacrawler

Permet d'obtenir rapidement des réponses issues de 7 moteurs importants (Lycos, Webcrawler, Excite, AltaVista, Yahoo, HotBot et Glaxy).

Savvysearch

Interroge 28 moteurs dont DéjàNews (groupe de news).

All-in-one

Méta-moteur français

Indiqué par la liste ADDBS.

liste des moteurs Internet

Liste exhaustive on-line des moteurs de recherche sur le web.

Permet de retrouver rapidement un moteur dont l'adresse n'a pas été intégrée au répertoire.

World-Wide Web Servers: Summary

Annuaire de sites par pays

LIENS VERS...

Arts

Fine arts library

Joconde est une base contenant 494 numérisations de peintures, écrits, sculptures et photographies provenant de 60 musées de France.

Bases ASCII

Derwent

Accès à Patent Explorer

Pharmaprojects

Disclosure

DAFSA

IDdb

Knight rider

Bases Web

Bases gratuites fournissant réellement une information pertinente.

MEDLINE

Version de l'Evaluated Medline : il s'agit de la forme la plus complète de Medline sur Internet accessible gratuitement. En effet, nous retrouvons l'intégralité des possibilités d'interrogation du Medline sur CD-ROM.

BIAM

Banque d'Information Automatisée sur le médicament : banque française d'information sur 2600 principes actifs présents dans les médicaments utilisés en France. Pour chaque spécialité, nous pouvons consulter les propriétés, indications, contre-indications et précautions d'emploi, les effets secondaires, la composition, la posologie, le traitement en cas de surdosage, les équivalents français et étrangers. Les données sont fournies par les laboratoires pharmaceutiques.

Brevets

Liste de sites associés

IBM patents

pages numérisées du brevet

Companie Nationale des Conseils en Propriété Industrielle

2 rubriques intéressantes :

liste alphabétique des consultants français

abécédaire de la propriété industrielle

OMPI (WIPO)

Organisation Mondiale de la Propriété Intellectuelle

World Intellectual property Organisation

Micropatent ; Official Gazette

Production hebdomadaire

office US des brevets et des marques

3\$US le brevet

Chimie

Conférences et meetings chimie

Chemical & Engineering News

1997-1998 Calendar of Meetings & Events

1997 - 2007 - ACS Calendar of Meetings

Modélisation moléculaire

Modélisation moléculaire

Liste de sites sur le sujet

Links to Biochemistry on the Web

Accès à partir du site Martindale's Chemistry Center, après avoir effectué une opération Find "molecular modeling".

Biochemistry Graphics on the Web

NIH center for molecular modeling

Le National Cancer Institute et le National Institute of Health possèdent des sites reliés entre-eux.

Tripos Homepage

IUCr - International Union of Crystallography

Le site fédérateur, le plus intéressant et gratuit.

Crystallographic tool

nécessite le chargement de 2 applications assistantes : VRML browser & PDB format

Data and Databases for crystallography

Chimie analytique

Mass Spectroscopy

Spectroscopy

liste de sites sur le serveur EMS (European Mass Spectrometry)

RMN

RMN Internet resources

Bruker USA

Bruker Germany

Service on line de m-à-j de programmes pour la MNR

visualisation de molécules en 3D

Chemical structures database

Library of 3-D Molecular Structures

MathMol Library

Department of Chemistry : Imperial College of Science

Permet d'accéder à de petites bases de données moléculaires, mais le nombre d'informations sur un produit est important (spectres, données physicochimiques). Le nombre de formats de fichiers étant très élevé, il faut autant d'applications assistantes!

Molecular Design Limited (MDL)

Page d'accueil du site MDL.

ChemWeb

Possibilité de changer toute la fiche de membre.

Chem links

IUPAC

Nomenclature des composés et enzymes, glossaires de termes, poids moléculaires.

Martindale's: Virtual Chemistry Center

Site majeur comportant de nombreux liens.

Attention, site de taille conséquente : 220K donc long à télécharger aux heures où le trafic est élevé.

Prous publications

Droit

Journal Officiel-Textes généraux

Permet la visualisation des textes généraux des 5 derniers numéros sous forme de documents scannerisés.

Economie et Bourse

Investir (journal)

Présentation des indices réactualisés très régulièrement (toute les 30 min), cotations (matin et soir), didacticiel sur le fonctionnement de la bourse. Création d'une liste personnelle de valeurs : une fois abonné, la consulter, la modifier, créer une nouvelle liste et recevoir chaque jour, automatiquement, les informations qui la concernent par mail. Plusieurs formats de réception sont disponibles (Métastock, EBP, Boursitel, excel/texte...) et toutes les valeurs traitées dans la rubrique cotations peuvent figurer dans la liste..

Fournisseurs et Entreprises

Bottin

Euronline

Utile pour identifier un fournisseur. Une limite cependant, les résultats sont

inégaux quant à la pertinence des réponses.

Europages

Site proche du précédent : Euronline.

Thomas register

Permet d'identifier un fournisseur américain.

EDGAR

Electronic Data Gathering, Analysis, and Retrieval system : permet d'obtenir les rapports annuels, rapports financiers périodiques et autres codes SEC.

Permet d'identifier les filiales d'un groupe.

Eudranet

Informations européennes sur l'industrie pharmaceutique

Agence européenne pour l'évaluation du médicament : www.eudra.org/emea.html

Neurobiologie

New York Academy of Sciences Home Page

Contient une liste de références classées par domaine. Possibilité de commander ces références avec un tarif préférentiel en étant membre de l'académie (NYAS).

The Journal of Neuroscience Online

Texte intégral des articles depuis Mai 96.

Accès gratuit.

Pointe vers le site de la Society For Neuroscience (SFN).

Journal of Neurochemistry

Donne accès à la table des matières et aux abstracts de la collection à compter de Juillet 96. Accès gratuit, mais pas de texte intégral.

Oncologie

Developmental Therapeutics Program, DCTDC, NCI

CancerWEB: Main Page

Pathologie et Médecine

Medscape

Classement par thèmes médicaux.

CHU de Rouen

Important site fédérateur français (consultable aussi en anglais et en italien).

Ressources Biomédicales sur l'Internet

Classement des ressources biomédicales par sujet.

Pharmacie

Conférences

Pharmacy Related Conferences

Pharmweb

Contient de nombreux liens vers des sites

miroirs.<http://pharmweb1.man.ac.uk/pwmirror/>

articles d'information sur Internet et la Pharmacie par D'Emanuele, A.

Martindale's "Virtual Centers"

Pharmaceutical Information Network

Permet d'accéder à un répertoire de médicaments internationaux : interrogation possible par nom de spécialité ou nom DCI.

Virtual Library - Pharmacy

Guide to Health Economics, Medical and Pharmacy Resources on The Net

Pharmacy-related Internet resources

Groupement provincial de l'industrie du médicament

Propose des liens intéressants vers d'autres sites notamment celui de la FDA (cf site suivant).

FDA**Phytologie**AGIS Home Page

Agricultural Genome Information Server :
base produite par the Genome Informatics Group at the National Agricultural
Library (NAL).

Phytochemical and Ethnobotanical Databases**Presse**Silverplatter medical journals

Ce site pointe sur 120 Journaux en ligne avec indication brève du contenu (texte
intégral ou simple résumé)

ElsevierINIST

Commande de documents à l'INIST.

National Library of Medicine & National Institute of Health

le communiqué de presse de la NLM

les données sur Medline

les 2 outils de recherche de la NLM : PubMed : les données les plus récentes
figurent dans la base (Premedline), et il existe des liens entre les références et sites
des éditeurs.

GratefulMed : aptitude à assister le praticien dans ces recherches.

AFP

Agence France Presse :

*donne une liste des journaux en ligne francophones.

*fait le point sur les nouvelles du jour dans le monde entier.

*le reste des dépêches est accessible via Globe On-line (payant).

ILAR

National Academy Press Bookstore (NAS)

psychological well-being of nonhuman primates

Science Online

Possibilité d'obtenir l'accès à toute la base et notamment le texte intégral des
articles en indiquant les 8 chiffres du numéro AAAS.

JAMA: The Journal of the American Medical AssociationLes EchosLa Tribune Desfossés

accès libre, texte intégral, archive depuis janvier 1996, bourse française (actualités
toutes les 30 min), rapports annuels de certaines entreprises.

Pharmaceutiques

Tous les articles ne sont pas disponible en ligne mais on peut interroger toutes les
archives depuis le n°1 jusqu'au numéro courant (n°48 le 01/07/97) sur des thèmes
proposés notamment fusion et acquisition

(veille). La partie 'chiffre' est aussi à consulter régulièrement pour être informé
des nouvelles concernant le monde pharmaceutique.

Journal du Téléphone

Information sur le monde de la communication mobile et Internet.

ZootechnieAAALAC

Association for Assessment and Accreditation of Laboratory Animal Care

DIVERS**Logiciels**

Chemistry MIME

logiciel de modélisation moléculaire et application assistante de navigateur.
EVANS T, Intranet conception et administration, 1996 p450, Ed Simon &
Schuster Macmillan.
<http://www.ssm.fr>

WWW Viewer Test Page

Cette page permet de visualiser la majorité des formats informatiques utilisés en chimie. Il est possible de tester si les applications assistantes sont en place grâce à des exemples fournis.

Windows Books

Liste de données écrites sur l'OS Windows

shareware**Sander's keyscreen previewer**

Offre un choix de shareware et freeware classés par domaine notamment l'Internet. Permet, avant tout téléchargement de fichier, de visionner l'écran principal et les fonctions majeures du logiciel.

Images et Cliparts**visible_gallery.html****Galerie d'images (Visible human)****the Clipart Directory****Free cliparts**

site de cliparts gratuits et liens vers des sites similaires

Welcome to Canada NewsWire's Portfolio Email**Fondation Kastler : Home page**

Le site donne des infos précises sur les modalités d'accueil en France de chercheurs étrangers.

Pesticide Information Profiles**Mars Pathfinder - Welcome to Mars!**

Galerie d'images et autres documents sur la mission Pathfinder.

Doc'INSA - INSA de Lyon

Le service Sapristi donne des informations sur l'internet et des moyens de recherche.

CRPF3's bookmark

BIBLIOGRAPHIE

1. ARCHIBAUD JL et AL
L'Internet professionnel : témoignage, expériences, conseils pratiques de la communauté enseignement et recherche
CNRS édition, 1995, 448p
2. LE GUELVOUT A.
Créer un service web de diffusion de l'information
Documentaliste-Science de l'information, 1997, vol 34, n°3, 179-183
3. PERRIAULT I.
Le métier de documentaliste et l'Internet
Archimag juillet/août 1997 n°106
4. URFIST Alsace Lorraine Franche-Comté
Chercher l'information sur Internet : Le défi [En ligne]
site Web : http://www-scd-ulp.u-strasbg.fr/urfist/recherche_information/sommair.htm
Page consultée le 5 août 1997
5. AYMONIN D
Les Nouvelles Technologies dans les bibliothèques
Edition du Cercle de la Librairie ; Nov 1996. - ISBN : 2-7654-0623-5
6. BENSOUSSAN A.
Internet : aspects juridiques
HERMES édition, 1996, 127p
7. BARTHELEMY P
La pratique quotidienne de l'Internet dans un laboratoire de recherche fondamentale
communication au colloque " La communication de l'IST dans l'enseignement supérieur et la recherche " - 17 mars 1995 - Bordeaux Talence, repris dans la Lettre de l'URFIST Midi Pyrénées, N°3 Mai juin 1995
8. FOENIX-RIOU B.
ECHO : un nouveau robot pour la francophonie
Netsource, n°6, janvier/février 1997.

9. DAUCHOT O.
Utilisation des méta-moteurs : avantages et inconvénients
Netsources, n°7, mars/avril 1997
10. LARDY J.P.
Les outils de recherche d'information sur Internet
Documentaliste, Sciences de l'information, 1996, vol.33, n°1, p33-37
11. PLOURDE J.N.
Bibliographie sur l'évaluation d'outils de recherche
site Web : <http://mistral.ere.umontreal.ca/~beaudryg/cursus/vol1no2/plourde.html>
Page consultée le 29 juillet 1997
12. URFIST Alsace Lorraine Franche-Comté
Internet et technique [En ligne]
site Web : http://www-scd-ulp.u-strasbg.fr/urfist/internet_technique/question.htm
Page consultée le 5 août 1997
13. LARCHER S.
Huit logiciels pour naviguer à moindre frais
Windows Plus, juillet/août 1997, n°44, 74-85
14. Société CYBION
Les Agents "semi-intelligents" [En ligne]
site Web : <http://www.cybion.fr/agents/semiagents.htm>
page consultée le 12 juin 1997
15. LE GUELVOUT A.
Créer un service web de diffusion de l'information
Documentaliste, Sciences de l'information, 1997, vol.34, n°3, p179-183
16. LEVY-ABEGNOLI T.
Le web prolonge le client-serveur
01 réseaux, mars 1996, n°24, P99-105
17. GUINIER D.
Autoroutes de l'information et Internet : sécurité des entreprises
Expertises des systèmes d'information, juillet-août 1995, p264-267

18. ASTRUC C.

Nouvelle architecture : l'informatique en danger

Réseaux et télécom, février 1997, p28-32

19. LORENZINI F.

Protéger son réseau avec un firewall

Informatique magazine, nov 1995, p28-32

20. BOUTIN Y.

Un site Web bien né

CD-RAMA, sept 1996, n°19, p40-45

21. URFIST Alsace Lorraine Franche-Comté

Chimie et Biologie sur Internet, panorama des ressources Juillet 97 [En ligne]

site Web : http://www-scd-ulp.u-strasbg.fr/urfist/internet_chimie_biologie/principa.htm

Page consultée le 5 août 1997.