

M 1998 ID ST 03

1419

**ENSSIB**

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

**Université  
Claude Bernard-Lyon 1**

**DESS en INFORMATIQUE DOCUMENTAIRE**

**RAPPORT de STAGE**

**INTERNET et LA BIOTECHNOLOGIE.**  
Conception et réalisation d'un site web en vue d'améliorer  
l'accès aux sources d'informations existantes.

**Cécile CHEVALLIER**

Sous la direction de  
Mme Véronique EMERAT

SANOFI RECHERCHE  
Centre de Labège

BIBLIOTHEQUE DE L'ENSSIB



8138997

**1998**

**ENSSIB**

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

**Université  
Claude Bernard-Lyon 1**



**DESS en INFORMATIQUE DOCUMENTAIRE**

**RAPPORT de STAGE**

**INTERNET et LA BIOTECHNOLOGIE.**  
Conception et réalisation d'un site web en vue d'améliorer  
l'accès aux sources d'informations existantes.

**Cécile CHEVALLIER**

Sous la direction de  
Mme Véronique EMERAT

SANOFI RECHERCHE  
Centre de Labège

**1998**

## REMERCIEMENTS

*Je remercie Max Cailleau pour son accueil au sein du centre de recherche de Sanofi à Labège.*

*Je tiens à remercier tout spécialement Véronique Emerat, maître de stage, de m'avoir permis de travailler sur ce sujet tellement intéressant et enrichissant. Je la remercie également pour sa confiance, tous ces conseils et son aide.*

*Je remercie également Jacqueline Troy, Claudette Olives de leur soutien et bonne humeur quotidienne, et Jean Noël Vallat pour nos discussions très instructives. Un double merci à Véronique et Jacqueline pour la relecture du mémoire !*

*Je n'oublie pas Stéphane et Christophe, informaticiens, pour leurs nombreuses interventions et leur disponibilité.*

**Titre :** Internet et la biotechnologie. Conception et réalisation d'un site web en vue d'améliorer l'accès aux sources d'informations existantes.

**Auteur :** Cécile Chevallier

**Résumé :** Internet est devenu une source de renseignements incontournable. Mais il continue de poser à l'utilisateur des difficultés dans la recherche de l'information pertinente. Ce rapport présente la conception et la réalisation d'un site web, interne à un centre de recherche, visant à rendre plus accessible l'information en biotechnologie.

**Mot-clés :** Biotechnologie - Internet - Site web - Recherche d'informations - Moteur de recherche - Utilisateur final- Accès information.

**Title :** Internet and biotechnology. Conception and realization of a web site in order to improve the access to the existing sources of information.

**Abstract :** Internet has become an unavoidable source of information. But it continues to present some difficulties to the user in his research of information. This report presents the conception and realization of a research center's web site who's aim is to make more accessible information in biotechnology.

**Keywords :** Biotechnology - Internet - Web site - Information retrieval - Search engine - End user - Information access.

## Table des matières

## Introduction.

<b>I. SANOFI.....</b>	<b>7</b>
A. SANOFI.....	7
B. SANOFI RECHERCHE.....	7
C. LE CENTRE DE LABÈGE.....	7
D. LE SERVICE D'INFORMATION DOCUMENTAIRE.....	8
1. <i>Les missions.</i> .....	8
2. <i>Le fonds documentaire.</i> .....	9
3. <i>Les outils de travail.</i> .....	9
<b>II. CONCEPTION DU SITE INTERNET.....</b>	<b>11</b>
A. LE CAHIER DES CHARGES.....	11
1. <i>Le projet.</i> .....	11
2. <i>Analyse de l'existant.</i> .....	11
a) L'environnement de l'accès à Internet.....	11
b) Le matériel.....	11
c) Le site d'accueil actuel.....	12
d) La population concernée.....	13
3. <i>Analyse des besoins.</i> .....	13
a) Le questionnaire.....	13
b) Analyse des résultats.....	14
(1) Fréquence de connexion des utilisateurs à Internet.....	14
(2) Langues utilisées sur Internet.....	15
(3) Domaines de recherche d'informations.....	15
(4) Types d'informations recherchés.....	17
(5) Page d'accueil actuelle.....	18
(6) Sélection des sites.....	18
(7) Les moteurs de recherches.....	19
(8) Mode d'interrogation des moteurs de recherche.....	20
(9) Sources d'informations recherchées.....	21
(10) Méthodologie de recherche sur Internet.....	21
(11) Profils personnalisés sur Internet.....	22
(12) La séance idéale sur Internet.....	22
c) Conclusions.....	24
(1) Profil d'utilisateurs.....	24
(2) Le site d'accueil.....	24
d) Les objectifs du site d'accueil.....	25
4. <i>Les différentes rubriques du site d'accueil.</i> .....	25
a) Les domaines.....	25
b) Les journaux en texte intégral.....	26
c) La méthodologie de recherche sur Internet.....	26
d) Envoyer nous un message !.....	27
B. CONCEPTION DE LA PARTIE RECHERCHE D'INFORMATIONS.....	27
1. <i>La recherche d'informations sur Internet.</i> .....	27
2. <i>Les moteurs de recherche.</i> .....	29
a) Les différents types de moteurs de recherches.....	29
(1) Les moteurs de type répertoire.....	29
(2) Les moteurs de type index.....	29
(3) Les méta-moteurs.....	30
b) Différentes approches.....	30
(1) Approche avec un seul moteur.....	30
(2) Approche avec les méta moteurs.....	30
c) Présentation de quelques moteurs disponibles sur Internet.....	31
(1) Caractéristiques des moteurs de recherche.....	31

(2) Langage d'interrogation des moteurs et méta-moteurs.....	31
(3) Méthodes de tri et de classement des résultats.....	32
d) Evaluation des moteurs de recherche et méta-moteurs de recherche par rapport aux besoins des utilisateurs.....	32
(1) Moteurs de type index.....	33
(2) Les moteurs de type répertoire.....	33
(3) Les méta moteurs.....	33
(4) Résultats des recherches.....	34
(5) Exploitation des résultats.....	38
(6) Conclusions.....	39
e) Description détaillée des caractéristiques des moteurs de recherche sélectionnés.....	39
(1) Les moteurs de recherche de type index.....	39
(2) Les moteurs de recherche de type index répertoire.....	44
(3) Le méta-moteur de recherche.....	45
(4) Conclusion quant à la démarche de recherche à adopter.....	46
3. Recherche par type d'informations.....	46
a) Objectif.....	46
b) Les différents types d'informations.....	47
(1) Sélection.....	47
(2) Les types d'informations.....	47
c) Sélection des liens et des moteurs spécialisés.....	48
d) L'équation de recherche et AltaVista.....	48
C. LA SÉLECTION DE SITES PAR DOMAINES.....	50
1. Objectif.....	50
2. Sélection des domaines.....	50
3. Sélection des sites.....	51
4. Liens vers la rubrique "Web Alert " des périodiques.....	51
D. LA CHARTE GRAPHIQUE.....	52
1. Les règles générales de développement.....	52
2. Présentation générale du site.....	52
3. L'accueil.....	52
4. Les journaux.....	53
5. Les domaines.....	54
6. La recherche d'informations.....	55
<b>III. REALISATION ET MISE EN OEUVRE.....</b>	<b>57</b>
A. LES MOYENS INFORMATIQUES.....	57
1. Le logiciel FrontPage.....	57
a) L'explorateur.....	57
b) L'éditeur.....	57
c) Le serveur Web personnel.....	58
d) Microsoft Image Composer.....	58
B. LA CRÉATION DU SITE.....	58
1. L'architecture du site.....	58
2. Les pages html.....	59
3. Des exemples de pages.....	59
C. LA CONFIGURATION DU SITE.....	60
D. LA GESTION DU SITE.....	60
1. La maintenance.....	60
a) Les lien hypertextes.....	60
b) Le serveur, les ordinateurs.....	61
2. La mise à jour.....	61
<b>IV. CONCLUSION.....</b>	<b>62</b>

BIBLIOGRAPHIE.

ANNEXES.

Internet est devenu un outil incontournable pour l'entreprise. Il ne remplace pas les sources classiques d'informations mais est complémentaire. Il présente de nombreux avantages : il est accessible par tous et facile d'utilisation. Il permet aussi de communiquer en temps réel et d'accéder beaucoup plus rapidement à la connaissance.

Une des originalités de ce « réseau des réseaux » est l'hétérogénéité de son contenu et de la structure des documents.

Cette particularité nécessite un intermédiaire auprès des utilisateurs : le documentaliste, ou cyber-documentaliste, met au service des usagers sa bonne connaissance des outils de recherche et des ressources d'information.

Ce rôle m'a été confié afin d'élaborer un site d'accueil pour faciliter l'accès aux ressources pertinentes disponibles sur Internet.

Le projet s'est réalisé au centre de recherche de Sanofi, situé à Labège, sous la tutelle de Mme Véronique Emerat, qui, outre sa fonction de chef de service du centre de documentation, est une cyber-documentaliste chevronnée.

Dans un premier temps une enquête a été menée auprès des utilisateurs du centre afin de mieux cerner leurs attentes, craintes et besoins par rapport à Internet. Dans un deuxième temps, une méthodologie de recherche d'informations et une sélection de ressources ont été conçues en adéquation avec les besoins des chercheurs. Enfin, le site d'accueil a été réalisé pour organiser, mettre en valeur et rendre accessible toutes ces informations.

## **I. SANOFI.**

### **A. SANOFI.**

Le groupe Sanofi, filiale d'Elf Aquitaine, a développé ses activités dans la santé (pharmacie, diagnostics et nutrition animale) et dans la beauté (Yves Saint Laurent, Van Cleef & Arpels, Roger Gallet).

Il est le deuxième laboratoire pharmaceutique français. Les axes thérapeutiques majeurs en santé humaine sont le cardiovasculaire, le système nerveux central, l'inflammation et l'ostéoporose, ainsi que la cancérologie.

### **B. SANOFI RECHERCHE.**

Sanofi Recherche dispose de neuf centres de recherche à travers le monde dont quatre en France, les autres étant situés en Espagne, Grande Bretagne, Hongrie, Italie et Etats-Unis. Dans ces centres se répartissent les trois étapes de l'élaboration des produits : la recherche, le développement pré-clinique et le développement clinique.

### **C. LE CENTRE DE LABEGE.**

Spécialisé dans le domaine des biotechnologies (utilisation à des fins industrielles des connaissances de la biologie), le centre de Labège se situe à la charnière entre les sciences fondamentales de la vie et leurs applications industrielles.

La recherche à Labège porte sur trois domaines : la santé humaine, la cosmétologie et le pétrole. Ses missions sont la production de protéines recombinantes, le clonage et l'identification de nouveaux récepteurs, la biologie des maladies (étude des phénomènes biologiques au niveau cellulaire), la maîtrise de la recherche appliquée et du développement pré-industriel en tenant compte de la réglementation, des coûts, des délais, ...

Le centre est consacré à deux étapes de l'élaboration des produits : la recherche amont et le développement pré-clinique:

- la recherche amont pour laquelle travaillent les unités suivantes :
  - unité de biologie moléculaire du gène.
  - unité de chimie organique.
  - unité de technologie des cellules animales recombinées.
  - unité de microbiologie.
  - unité de biochimie des protéines.
  - unités des substances naturelles.
  
- le développement pré-clinique concerne :
  - l'unité pilote biotechnologie, qui permet la production de quantités importantes de principe actif et l'évaluation des coûts.
  - l'unité analyse et contrôle qualité.

## **D. LE SERVICE D'INFORMATION DOCUMENTAIRE.**

### **1. Les missions.**

Une des premières missions est l'accueil du public, accueil assuré par le personnel de la documentation.

Le service d'Information Documentaire (SID) a pour but de rechercher, diffuser et mettre à disposition auprès de l'ensemble du personnel du centre de Labège toute l'information scientifique et technique nécessaire à son activité. En tant que prestataire de services, ses missions sont les suivantes :

- opérations courantes d'un centre de documentation : commandes, réception, facturation, bulletinage, catalogage, indexation et archivage,
- recherches bibliographiques sur bases de données,
- mise à disposition des documents primaires,
- valorisation du fonds documentaire par :
  - l'élaboration de différents produits documentaires : liste d'acquisitions, profils informatisés issus des bases de données.

- la promotion de produits en libre-service.

- formation des utilisateurs aux produits en libre service et accessibles par le réseau,
- conception et développement de nouvelles applications pour permettre une exploitation optimale du fonds et des outils dont dispose le SID.

## **2. Le fonds documentaire.**

- Les ouvrages : 3500 exemplaires sont destinés à l'ensemble des services et sont divisés en ouvrages généraux (encyclopédies, annuaires, ...) et en ouvrages spécialisés.
- Les périodiques : le centre possède une collection de 265 périodiques dont 215 sont vivants et proposés à la consultation.
- Les cédéroms : de bases de données (Medline, Current Contents, Biotechnology Abstracts), d'ouvrages encyclopédiques (Encyclopédie Universalis, Dictionary of Natural Products, la Pharmacopée européenne, ...) et des catalogues (par exemple celui de l'ATCC : American Type Culture Collection). Ils sont consultables grâce à deux postes en libre accès au SID et sur les postes d'utilisateurs pour ceux qui en ont fait la demande.
- Les rapports : les thèses ainsi que les rapports de stage du centre sont consultables au SID.

## **3. Les outils de travail.**

- La gestion :

Le gestion des ouvrages et des périodiques est faite avec Gedbib, application du logiciel documentaire BASIS+ en mode client/serveur sous Windows. Cette base est unique et commune à l'ensemble des services d'information documentaire de Sanofi, pour qui elle a été spécifiquement développée. Elle permet toutes les opérations de gestion du fonds documentaire (catalogage, bulletinage,

interrogation, ...). Une autre application appelée Gedpub, intégrée dans Gedbib, gère les publications du groupe et celle des chercheurs, ainsi que les rapports.

- L'accès à Internet :

Actuellement, l'accès se fait par un PC isolé, en libre service, avec connexion par modem.

## **II. CONCEPTION DU SITE INTERNET.**

### **A. LE CAHIER DES CHARGES.**

#### **1. Le projet.**

Le site d'accueil spécifique au centre de recherche de Sanofi Labège doit répondre aux objectifs suivants : rendre plus accessible l'information disponible sur Internet, faciliter la recherche d'informations pertinentes tout en réduisant le temps de connexion.

#### **2. Analyse de l'existant.**

##### a) L'environnement de l'accès à Internet.

Une décision stratégique a été récemment prise, au niveau de Sanofi Recherche, par les représentants des services informatiques et scientifiques. L'accès à Internet se fera *via* le réseau interne de Sanofi par deux postes dédiés, en libre service au centre de documentation.

##### b) Le matériel.

- L'accès Internet.

Actuellement, il existe un seul poste connecté à Internet, poste en accès contrôlé et mis en service depuis un an. Ce poste est isolé du réseau, équipé d'un modem (33600 bps/min) ; la liaison est établie par le prestataire de service, WorldNet.

Dans quelques semaines, ce poste sera remplacé par deux postes connectés à Internet, avec un accès réseau, sous système d'exploitation Windows NT. L'accès à Internet par ce biais, consiste en une connexion par le réseau local de type Ethernet, puis passage par les deux « coupes feux » de Sanofi pour déboucher sur le web par une liaison Transpac. L'accès à Internet est soumis à autorisation hiérarchique avec prise de connaissance des directives de sécurité de Sanofi.

- Les caractéristiques des postes de travail sont :
  - poste dédié, en libre service.
  - PC 575, 16 Mo RAM, 500 Mo disque (configuration minimale).
  - Windows NT 4.0.
  - Navigateur Netscape 4.04.
  - Anti-virus : Norton de Symantec.
  - Connexion au réseau local.

c) Le site d'accueil actuel.

Il existe un site d'accueil mis en place en même temps que l'accès Internet. Ce site, présent en page d'accueil, est une copie des signets de Sanofi. Les adresses répertoriées sont accompagnées d'un descriptif succinct. Elles sont regroupées en sous-thèmes, eux mêmes regroupés en thèmes généraux.

Les thèmes sont les suivants :

<b>Thèmes</b>	<b>Types d'adresses recensées</b>
Cardiology	Coronary, general, resources
Central Nervous System	Journals, resources
Internal medicine	Endocrinology, inflammation, osteoporosis
Oncology	Journals, resources
Genetics	Databases, resources
Clinical trials	Resources
Pharmaceutical competitors	Companies
Pharmaceutical markets	Epidemiology, general, resources
Healyh authorities	Resources
Beauty	Companies
Chemistry	Companies, databases, journals, resources
Diagnostics	N'existe plus

Patents	Databases, resources
Journal contents	Journals, magazines, scientific titles
Administration	Authorities, legislation
Computer technology	Magazines, sharewares
Internet	Guides, search engines
Finance	Companies, general

Chaque sous thèmes pointent vers cinq adresses au maximum, essentiellement de sites ressources.

d) La population concernée.

Selon les recommandations du service Informatique, tout membre du site de Labège a la possibilité de se connecter. L'autorisation de l'ouverture d'un compte est soumise à approbation hiérarchique.

Le public actuel est composé en grande majorité de chercheurs.

Il y a actuellement 46 comptes utilisateurs. Leur profil est indiqué dans l'analyse des besoins, question numéro 1.

### 3. Analyse des besoins.

a) Le questionnaire.

Les objectifs de ce questionnaire sont multiples. Tout d'abord, le SID avait besoin d'un retour concernant le système mis en place il y a un an. Ensuite, il y avait une demande de la part de Mme Emerat d'optimiser l'accès aux ressources d'Internet par l'élaboration d'un site d'accueil et d'une évaluation des outils de recherche présents sur Internet. Pour faire cela, il fallait donc en premier lieu, évaluer les besoins et les attentes des utilisateurs par rapport à Internet.

J'ai contacté les utilisateurs ayant un compte internet afin de prendre rendez vous avec eux.

J'ai choisi de les rencontrer personnellement afin d'établir un premier contact et d'engager un dialogue.

Les résultats du questionnaire ont été traités avec le logiciel Excel, très pratique pour calculer les pourcentages et faire des graphiques explicites. Un code est donné à chaque possibilité de réponse. Le nombre d'occurrence des codes est calculé, ainsi que leurs fréquences. Des graphiques sont établis à partir de ces fréquences.

b) Analyse des résultats.

Sur le centre de recherche de Labège, 46 personnes possèdent un compte Internet et 36 utilisateurs ont répondu au questionnaire. Ceci donne un taux de réponse d'environ 78 %. Les vingt deux autres pour cent sont des personnes qui n'ont pas voulu répondre, des personnes qui ont pris un compte pour consulter à la place de quelqu'un de leur service et des personnes que je n'ai pas réussi à joindre.

Le taux de réponse obtenu permet de souligner que les résultats sont significatifs, exploitables et qu'ils reflètent bien l'ensemble des besoins des utilisateurs.

(1) *Fréquence de connexion des utilisateurs à Internet.*

<p>A quelle fréquence vous connectez vous ?</p> <ul style="list-style-type: none"> <li>• de temps en temps</li> <li>• 1 fois/mois</li> <li>• 1 fois/semaine</li> <li>• 1 fois/jour</li> </ul>	
---	--

Le résultat graphique est en annexe page 8.

53% des utilisateurs se connectent de temps en temps : ceci est dû à un manque de disponibilité des utilisateurs, un manque de maîtrise, ou juste au fait qu'ils ne prennent pas le temps de faire leurs recherches.

6% se connectent 1 fois/mois et 26% se connectent au moins une fois par semaine.

15% des utilisateurs disent se connecter une fois par jour.

Beaucoup de chercheurs m'ont fait part de leur intention, à court terme, de se connecter deux à trois fois par semaine !

(2) *Langues utilisées sur Internet.*

<p>Quelle langue utilisez vous principalement pour :</p> <ul style="list-style-type: none"> <li>• la consultation : <ul style="list-style-type: none"> <li>- français</li> <li>- anglais</li> <li>- les deux</li> </ul> </li> <li>• la recherche d'informations: <ul style="list-style-type: none"> <li>- français</li> <li>- anglais</li> <li>- les deux</li> </ul> </li> </ul>	
--	--

Le résultat graphique est en annexe page 8.

En majorité, les utilisateurs utilisent l'anglais mais préféreraient que ce soit en français.

L'utilisation de l'anglais n'est un frein ni à la consultation ni à la recherche d'informations.

(3) *Domaines de recherche d'informations.*

	OUI
<p>Dans le cadre de vos activités, dans quel(s) domaine(s) cherchez vous de l'information ?</p> <ul style="list-style-type: none"> <li>• chimie</li> <li>• biochimie</li> <li>• biologie moléculaire</li> <li>• microbiologie</li> <li>• biotechnologie</li> </ul>	

<ul style="list-style-type: none"> <li>• immunologie</li> <li>• pharmacologie</li> <li>• toxicologie</li> <li>• neurobiologie</li> <li>• pathologie humaine</li> <li>• analyse</li> <li>• techniques de laboratoires / protocoles</li> <li>• documentation</li> <li>• informatique</li> <li>• sécurité</li> <li>• développement clinique</li> <li>• affaires réglementaires</li> <li>• assurance qualité</li> <li>• autres</li> </ul>	
---	--

Le résultat graphique est en annexe page 10.

Les domaines pour lesquels il y a le plus de besoins en information sont : la biotechnologie (14%), la biologie moléculaire (11%), la documentation (9%) plus particulièrement les journaux, la biochimie (8%), les protocoles (7 %) ou la microbiologie (7%).

En ce qui concerne les domaines de la réglementation et de l'assurance qualité, certaines personnes sont très intéressées mais elles sont moins représentées dans la "population" des utilisateurs, d'où de faibles pourcentages.

Pour ce qui est de l'informatique, ce sont surtout des informations ou des publications relatives aux logiciels utilisés dans le centre qui sont recherchées.

Enfin, la rubrique "Autres" regroupe la biosécurité, la biologie cellulaire, l'ethnobotanique, la botanique, la médecine populaire, les substances naturelles, et des textes de bases en réglementation.

(4) *Types d'informations recherchés.*

	OUI
<p>Par rapport aux domaines précédemment cités, de quels types d'informations avez vous besoin :</p> <ul style="list-style-type: none"> <li>• sites de vulgarisation</li> <li>• généralités sur un domaine (ex : cours, ...)</li> <li>• articles de journaux presse généraliste</li> <li>• articles de journaux presse spécialisée</li> <li>• brevets (ex : EPO, IBM Patent Server)</li> <li>• sites ressources (ex : Pasteur)</li> <li>• colloques, congrès, salons professionnels,...</li> <li>• formation continue</li> <li>• base de données : <ul style="list-style-type: none"> <li>séquences de gènes</li> <li>séquences de protéines</li> <li>données cristallographiques</li> <li>composés chimiques</li> <li>références bibliographiques</li> <li>nomenclature</li> <li>autres</li> </ul> </li> <li>• adresses électroniques</li> <li>• sites de fournisseurs</li> </ul>	

Le résultat graphique est en annexe page 11.

Les types d'informations les plus demandés sont l'accès aux bases de données (16%), les sites de fournisseurs et leurs catalogues (anticorps, produits chimiques) (14%), les articles de la presse spécialisée (12%), l'accès aux annonces de salons, congrès (11%) et l'accès à des adresses postales et électroniques (10,5%).

Le faible taux concernant les brevets provient peut être du fait qu'il existe un service Brevet au sein du centre de recherche et que la recherche sur ce type d'information est difficile.

Concernant les bases de données (résultat en annexe page 12), les utilisateurs souhaiteraient avoir accès à :

- des séquences de gènes : 24%
- des séquences de protéines : 22%
- des données cristallographiques : 7%
- des composés chimiques : 11%
- des références bibliographiques : 24%
- la nomenclature : 9% ( protéines, botanique : taxonomie et systématique)
- autres : 4% (localisation de gènes et sécurité)

(5) *Page d'accueil actuelle.*

	OUI	NON
Utilisez vous les adresses de la page d'accueil actuelle ?		
Remarques :		

Le résultat graphique est en annexe page 13.

La page d'accueil actuelle est consultée à 40%.

Ceci est dû à ce qu'il n'y a pas toutes les entrées et qu'elle n'a pas évolué depuis le début de sa mise en place. Elle est utilisée pour les adresses de moteurs de recherche principalement.

(6) *Sélection des sites.*

	OUI	NON
Seriez vous d'accord pour collaborer à la sélection des sites, en rapport avec vos activités, destinée à l'élaboration du site d'accueil ?		

Le résultat graphique est en annexe page 13.

Les utilisateurs sont prêts à 61% à participer à la sélection des sites en fonction de leurs activités : ce pourcentage prouve leur grand intérêt!

(7) *Les moteurs de recherches.*

	OUI	NON
<p>Avez vous déjà recherché des informations par des moteurs de recherche ?</p> <p>Si oui, lesquels ?</p>		
<p>Etes vous satisfait des réponses obtenues par les moteurs précédemment cités ?</p> <p>Remarques :</p>		
<p>Préférez vous :</p> <ul style="list-style-type: none"> <li>• les moteurs thématiques (type Yahoo) ?</li> </ul> <p>Lesquels ?</p> <ul style="list-style-type: none"> <li>• les moteurs généralistes (type Alta Vista) ?</li> </ul> <p>Lesquels ?</p> <ul style="list-style-type: none"> <li>• les méta-moteurs ?</li> </ul> <p>Lesquels ?</p>		

Les résultats graphiques sont en annexes pages 14 et 15.

Les moteurs de recherche sont utilisés par 70% des usagers, avec comme grands vainqueurs AltaVista et Yahoo. Sur ces 70% d'internautes, 53% ne sont pas satisfaits des réponses obtenus : il y a souvent trop de pages et une faible pertinence.

68% des utilisateurs préfèrent les moteurs de type généraliste et 26 % les moteurs de type annuaire.

Le faible pourcentage pour les méta-moteurs vient du fait qu'ils ne savent pas ce que c'est.

(8) *Mode d'interrogation des moteurs de recherche.*

	OUI	NON
Lorsque vous interrogez, est ce par : <ul style="list-style-type: none"> <li>• mots (sans autres précisions)?</li> <li>• mots dans les URLs ?</li> <li>• mots dans les titres ?</li> <li>• langage naturel ?</li> </ul>		
Lorsque vous interrogez, est ce par : <ul style="list-style-type: none"> <li>• mots-clés ?</li> <li>• concept ?</li> <li>• les deux ?</li> </ul>		

Les résultats graphiques sont en annexe page 16.

Les utilisateurs interrogent, à 80%, les moteurs de recherche sans tenir compte de la syntaxe précise car peu maîtrisent le langage d'interrogation particulier à chaque moteur.

Moins de 10 % des utilisateurs spécifient qu'ils cherchent les mots dans les titres et dans les adresses.

61% des utilisateurs interrogent par des mots-clés et 9% par des termes beaucoup plus généraux (exemple : biologie moléculaire). 29% interrogent avec les deux types.

En général, les chercheurs savent quelles informations ils veulent et essaient de passer le moins de temps possible connecté.

## (9) Sources d'informations recherchées.

	OUI	NON
Quelles sources d'infos vous intéressent ? <ul style="list-style-type: none"> <li>• sites Web</li> <li>• les listes de discussion</li> <li>• les deux</li> </ul>		

Le résultat graphique est en annexe, page 17.

Environ 55% des utilisateurs rechercheraient de l'information principalement dans les sites web. Les 45 autres pour cent sont intéressés par les deux types de sources d'informations.

Les 55% s'expliquent par le fait qu'ils sont déjà abonnés à des listes de discussion ou qu'ils ne veulent pas s'y abonner par manque de temps pour lire les messages.

## (10) Méthodologie de recherche sur Internet.

	OUI	NON
Seriez vous intéressé par une méthodologie de recherche sur Internet ? Si oui : <ul style="list-style-type: none"> <li>• en ligne ?</li> <li>• sur guide papier ?</li> <li>• les deux ?</li> </ul>		

Le résultat graphique est en annexe page 18.

94% des utilisateurs souhaitent une méthodologie de recherche sur Internet. Ils veulent pouvoir obtenir des réponses plus pertinentes et ne pas perdre de temps.

Plus de 50% d'entre eux souhaitent avoir accès à cette méthodologie en ligne et sur papier. La forme papier leur permettrait de pouvoir préparer leur interrogation et d'annoter le guide.

(11) *Profils personnalisés sur Internet.*

	OUI	NON
A terme, seriez vous intéressé par la mise en place de profils personnalisés sur Internet?		

Le résultat graphique est en annexe page 19.

79% des utilisateurs souhaitent la mise en place de profils personnalisés sur Internet.

(12) *La séance idéale sur Internet.*

Ce qui suit est une liste « brute » des remarques recueillies auprès des utilisateurs. Ces remarques reflètent les impressions des utilisateurs. En travaillant chaque jour à côté du poste internet, il s'avère qu'elles ne sont pas toutes fondées.

(a) *Au niveau du matériel.*

- Une grande majorité des utilisateurs souhaiteraient vivement avoir accès à Internet sur leur poste. Beaucoup ne se connectent pas car il faut venir au SID et qu'il faut attendre que le poste se libère.
- Ils souhaiteraient, si l'accès ne se fait pas sur leur poste, un ou deux postes de consultation en plus, avec un accès sur le réseau pour transférer leurs fichiers de données.
- Ils souhaiteraient des lignes à plus hauts débits.
- Une souris qui fonctionne.
- Un accès sécurisé. Certains n'interrogent pas Internet de peur que quelqu'un de l'extérieur ait accès à leurs recherches.

- Etre informé en cas de problèmes d'accès à Internet.
- Pas de déconnexions sauvages (problème lié au prestataire de service et/ou au modem).
- Une imprimante pour les images (couleurs et bonne résolution).
- Plus de mémoire pour récupérer les images.

*(b) Au niveau de l'accompagnement humain.*

- Une aide technique.
- Possibilités de recherche à la demande.
- Une personne à qui s'adresser en cas de difficultés lors des recherches.
- Une formation à la syntaxe de chaque moteur.
- Une connaissance des intérêts et inconvénients de chaque moteur.
- Des possibilités de formation personnalisée et approfondie.

*(c) Au niveau du site d'accueil.*

- Une méthodologie.
- Un guide technique.
- Convivialité.
- Un site vivant.
- Une pré-sélection d'adresses électroniques.
- Possibilité de pages personnalisées.
- Une sélection des meilleurs sites dans les différentes disciplines.
- Très basique pour démarrer.
- Ne pas perdre trop de temps pour démarrer : permettre une navigation simple.
- Avoir un moteur de recherche en page d'accueil.
- Un système de post it pour que ceux qui ont trouvé des sites intéressants puissent en faire bénéficier les autres.
- Avoir tous les textes réglementaires.
- Savoir rapidement ce qu'il y a dans le site d'accueil.

(d) *Divers.*

- Avoir un budget pour des articles en texte intégral au coup par coup (pour avoir certains articles avant la parution).
- Connaître les journaux présents online : ceux qui sont gratuits, ceux pour lesquels on a accès grâce au papier ainsi que les journaux électroniques.

c) *Conclusions.*

(1) *Profil d'utilisateurs.*

41% des utilisateurs disent se connecter une fois par jour à 1 ou plusieurs fois par semaine. D'autre part, comme il a déjà été précisé, un bon nombre d'internautes souhaitent augmenter leur fréquence de connexion. Il va donc y avoir une forte hausse de connexions et donc de demandes de disponibilité de matériel, de performances et de services.

Il apparaît aussi qu'il y ait un besoin de formation à plusieurs niveaux : une aide ponctuelle, en cas de difficultés, une aide et/ou formation personnalisée pour les utilisateurs en réelle difficulté, ainsi que des formations personnalisées et approfondies pour les plus avertis.

Les utilisateurs souhaitent aussi que cette personne soit à l'écoute de leurs besoins et possède une connaissance globale d'Internet pour les aiguiller, leur présenter des nouveautés et faire des recherches à la demande.

(2) *Le site d'accueil.*

Il existe un réel besoin d'informations de la part de ces utilisateurs, qui souhaitent accéder à ce qu'ils cherchent sans perte de temps.

Le questionnaire a permis de mettre en valeur la diversité des domaines et des types d'informations recherchés. Ils savent qu'ils ne peuvent ni tout connaître, ni faire une recherche exhaustive. D'autre part, un très grand nombre d'utilisateurs interrogent les moteurs de recherche à chaque connexion. Le fort pourcentage de réponses positives concernant l'élaboration d'une méthodologie (94%) ainsi que le taux d'insatisfaction (53%) par rapport

aux résultats obtenus, montrent qu'il y a aussi un énorme besoin d'aide à la recherche. Que cela passe par une meilleure connaissance de l'offre des outils de recherche (mise en évidence par le faible pourcentage d'utilisation des méta-moteurs et l'interrogation massive de seulement deux moteurs de recherche, Yahoo et AltaVista ) et des spécificités liées à chaque moteur (spécificités en terme de sources d'informations et de syntaxe pour l'interrogation, intérêts et inconvénients).

d) Les objectifs du site d'accueil.

Le site d'accueil qui va être mis en place doit permettre de résoudre tous ces besoins avec comme objectif de **trouver l'information pertinente en diminuant le temps passé sur Internet.**

Pour cela, une pré-sélection d'adresses dans chaque domaine et pour chaque type d'informations sera faite. Une méthodologie claire, concise, facile à utiliser et proposant un panel, pas trop important mais représentatif, d'outils de recherche. Enfin, le site doit être convivial et vivant.

#### 4. Les différentes rubriques du site d'accueil.

a) Les domaines.

Les domaines ont été choisis en fonction des activités du centre et du résultat de l'enquête.

- Biochimie
- Biologie Moléculaire
- Bookmark de Sanofi-Recherche (lien vers les signets de l'intranet de Sanofi)
- Documentation
- Législation/Réglementation
- Médecine - Santé
- Microbiologie
- Neurobiologie

- Protocoles/ Techniques de laboratoires
- Sécurité
- Substances Naturelles

Pour chacun des thèmes précédemment cités, un découpage sera effectué suivant le type d'informations: des bases de données, des sites ressources, des journaux, associations, ...

Sur la page d'accueil, seront en premier plan ce qui pourrait être le plus utilisé : le lien vers la page regroupant les journaux en texte intégral, le lien vers la page des moteurs de recherche ainsi que la page de méthodologie, un lien vers l'Intranet et les directives de sécurité.

b) Les journaux en texte intégral.

Cette page présente la liste des titres de périodiques présents en texte intégral sur internet et pour lesquels le SID a souscrit un abonnement online en supplément du papier.

c) La méthodologie de recherche sur Internet.

La méthodologie se fera selon deux approches.

La première approche se fera par types d'informations. Suivant le type d'information recherché, une sélection de sites, les moteurs à utiliser (avec un lien dynamique) et une équation modèle seront présentés. Cette dernière permettra à l'utilisateur de faire un copier - coller de la question sur le moteur de recherche. Ensuite, il lui suffira d'ajouter les termes spécifiques.

Les types d'informations sélectionnés sont : les brevets, conférences, journaux électroniques, thèses, articles, adresses électroniques, ... Il y aura aussi un lien vers les domaines concernés, si besoin est.

La deuxième approche portera plus sur une liste de moteurs de recherches, classés par types : annuaire, généraliste et méta-moteur. Pour chaque moteur, la syntaxe, la troncature, les spécificités ainsi que les intérêts et inconvénients seront présentés.

d) Envoyer nous un message !

Ce système offre la possibilité aux utilisateurs de faire part à l'administrateur d'adresses de sites qu'ils ont trouvés intéressants. Il permet aussi de demander une aide technique, une formation ou des recherches à la personne responsable quand elle n'est pas présente sur place.

## **B. CONCEPTION DE LA PARTIE RECHERCHE D'INFORMATIONS.**

La recherche d'information sur Internet est difficile. En effet, Internet est un système non limité et en évolution permanente. D'autre part, il n'y a pas de référence ni de contrôle, l'identification des sources est difficile et la pérennité de l'information pose un problème. Enfin, il existe pléthore de moteurs de recherches, chacun ayant leurs spécificités.

Les utilisateurs du centre de recherche ont fait part de leur intérêt pour l'élaboration d'une méthodologie de recherche.

Comme nous l'avons décrit précédemment dans le cahier des charges, une partie sera consacrée aux moteurs de recherche et une autre sera une sorte de mode d'emploi pour une recherche par types d'informations. Mais avant de se lancer dans cet ouvrage, nous ferons état des problèmes liés à la recherche d'informations sur Internet.

### **1. La recherche d'informations sur Internet.**

Rechercher une information sur Internet devient beaucoup plus complexe que de rechercher sur des bases de données traditionnelles. Afin de comprendre cela, il est nécessaire de connaître les caractéristiques de "l'information virtuelle" (Charron, 1997 ; Revelli, 1998) et de les mettre en relation avec les fonctionnalités des outils de recherche :

- *Une grande hétérogénéité dans le contenu:* l'information n'est ni contrôlée ni centralisée, et n'importe qui peut publier n'importe quoi sur Internet. D'autre part, il existe de nombreux formats de documents (postscript, pdf, ...) qui ne sont pas traités par les moteurs de recherche.

- *La localisation des documents* : l'information n'est pas centralisée et est sans cesse en mouvement. Les moteurs de recherche pallient ces problèmes par leur fréquence de visite des sites et une actualisation des liens.
- *La fragmentation* : un document sur Internet est fragmenté en différentes pages. Certains moteurs essaient d'indexer tout le site ou toutes les pages, mais en sont ils vraiment capables ? D'autres ne prennent qu'un échantillon de pages.
- *Le multilinguisme* : bien que ce soit un aspect important de l'Internet, aucun système de recherche n'est capable de le gérer. Ceci implique que toutes les requêtes, ou presque, sont en anglais et que la majorité des sites consultables sont également dans cette langue. Une grande quantité d'informations peut être perdue à ce moment là.
- *Le renouvellement continu* : la mise à jour plus ou moins fréquente des index permet de pallier cet aspect. Par contre, aucun outil n'ayant développé un système de nouveautés, il est impossible de pouvoir les reconnaître lors d'une interrogation.
- *L'accès aux informations* : de nombreuses données gratuites sont disponibles mais il existe aussi beaucoup d'informations à accès restreint.
- *L'information n'est pas structurée* : l'édition des documents n'est soumise à aucune norme ni standard.

En ce qui concerne les moteurs de recherche, nous savons depuis quelques temps qu'ils n'indexent pas toutes les pages du Web et que leurs bases ont des tailles variables (Lawrence et Gilles, 1998). Ceci est dû à des contraintes physiques (problèmes avec les cadres, les légendes d'images, ...) et des contraintes purement matérielles (la place nécessaire au stockage de toute cette information et la puissance des robots nécessaire à ce travail). D'autre part, le volume de pages de l'Internet est tellement important qu'il faut des semaines pour "en faire le tour".

Tous ces aspects mettent bien en valeur que **toute l'information disponible n'est pas accessible** et qu'il faut **une connaissance approfondie de celle-ci ainsi que des outils de recherche** présents sur l'internet avant d'effectuer une recherche documentaire.

## 2. Les moteurs de recherche.

### a) Les différents types de moteurs de recherches.

Il existe différents types de moteur de recherche (Habib et Balliot, 1997 ; Plourde, 1995) : le type répertoire, le type index et les méta-moteurs.

#### (1) *Les moteurs de type répertoire.*

L'information est catégorisée. Ce type est associé à tous les outils de recherche dont les fonctions de repérage reposent sur une classification qui organise l'information. La classification est effectuée par des personnes, ce qui implique une valeur ajoutée. Certains de ces moteurs possèdent une zone de recherche où l'on peut spécifier les mots que l'on cherche, mais leurs possibilités au niveau des opérateurs de recherche est très limitée. L'inconvénient de ce type de moteur est son manque d'exhaustivité. Ceci est dû à ce que l'augmentation du nombre de sites se fait par inscription des auteurs et que l'indexation humaine prend du temps. Ces moteurs sont à utiliser comme point de départ pour toute recherche sur le Web, et comme un outil de base pour la veille sur un thème.

Exemple : Yahoo, Magellan, Nomade.

#### (2) *Les moteurs de type index.*

L'information n'est pas compartimentée en catégories. Ce type de moteur repose sur des techniques automatisées pour répertorier l'information sur le Web : le texte de la page est partiellement indexé selon les moteurs.

La recherche s'effectue avec des mots-clés dans l'index. Ce système se rapproche du modèle des OPACs traditionnels mais le vocabulaire n'est pas contrôlé. L'avantage est que la mise à jour est fréquente et que les bases sont importants. L'inconvénient majeur est que l'indexation est automatique donc l'utilisateur peut avoir du mal à trouver ce qu'il cherche.

Ce type de moteur est à utiliser pour des recherches précises ou sur des thèmes pointus.

Exemple : AltaVista, HotBot, Excite, Infoseek.

*(3) Les méta-moteurs.*

Ce sont des logiciels qui font appel à plusieurs moteurs de recherche différents en parallèle pour effectuer les recherches (Liu, 1998). Ils essayent de faire correspondre la syntaxe de la recherche aux syntaxes spécifiques à chaque moteur, enlèvent les doublons et présentent une liste d'adresses de sites. Il est important de connaître les moteurs appelés lors de l'utilisation des méta-moteurs. Il faut aussi savoir que les résultats peuvent varier d'une recherche à une autre car les moteurs appelés ne sont pas toujours disponibles et que l'on obtient moins d'adresse puisqu'il ne prend que les 10 à 15 meilleurs réponses de chaque moteur.

Exemple : Metacrawler, Profusion, SavySearch, AskJeeves.

## b) Différentes approches.

Connaître les avantages et les inconvénients de chaque approche est indispensable (Adit, 1998 ; Hock, 1998 ; Notess, 1997).

*(1) Approche avec un seul moteur.*

- Avantages : on peut utiliser l'interrogation par phrases, les opérateurs booléens et faire des recherches dans des champs spécifiques. Les fonctionnalités du moteur sont exploitées à fond.
- Inconvénients : chaque moteur n'indexe qu'une partie de l'Internet et il faut connaître la syntaxe spécifique à chaque moteur de recherche.

*(2) Approche avec les méta moteurs.*

- Avantages :
  - envoie la requête à plusieurs moteurs en parallèle,
  - enlève les doublons,
  - trie les résultats par hôte, par mots-clés, par date,
  - deux types : les méta-moteurs online, que l'on va interroger sur internet, et ceux en local, c'est à dire ceux pour lequel le logiciel est installé sur le poste de travail ou sur un serveur local.

- Inconvénients :

- ne prend que les 10 à 15 premiers résultats de chaque moteur, il y a donc moins de résultats qu'avec un seul moteur de recherche,
- ne supporte pas bien les recherches booléennes,
- pas de possibilité de limiter sa recherche à des champs spécifiques.

c) Présentation de quelques moteurs disponibles sur Internet .

Il existe de nombreux moteurs de recherche disponibles. J'ai fait une pré-sélection parmi les plus connus. Ils sont présentés ci-dessous (Charron, 1997 ; Feldman, 1998 ; Notess, 1997 et les sites de différents moteurs).

(1) *Caractéristiques des moteurs de recherche.*

Les caractéristiques sont présentées en annexe pages 21 à 24.

(2) *Langage d'interrogation des moteurs et méta-moteurs.*

Le langage d'interrogation est développé en annexe page 25.

Remarques :

- Metacrawler :

- il ne faut pas d'espace entre les signes +/- et les termes de la recherche.
- un espace est nécessaire entre les termes.
- tous les moteurs de recherche appelés ne supportent pas ce type d'opérateurs.

- DogPile :

- l'opérateur NEAR est substitué en AND pour les moteurs qui ne le supportent pas.
- l'opérateur NOT est supprimé pour les moteurs qui ne le supportent pas.
- l'opérateur par défaut est AND.

- supporte les guillemets et les parenthèses mais sont enlevés si les moteurs ne les supportent pas.

(3) *Méthodes de tri et de classement des résultats.*

Chaque moteur de recherche possède ses propres algorithmes pour trier et classer les résultats.

Quelques critères pris en compte sont présentés ci-dessous (Hassani, 1998) :

- nombre d'occurrence de mots-clés extraits de la question dans le document trouvé.
- la valeur discriminante du terme : plus il est fréquent dans l'index, moins il aura d'importance.
- la proximité sémantique entre les termes recherchés dans le document.
- la position du mot recherché dans la structure logique du document.
- la syntaxe de la requête et l'opérateur OU : si les termes reliés par cet opérateur sont tous présents dans un document, celui-ci apparaîtra en premier.
- la pondération des mots : chaque moteur a son propre algorithme.

d) Evaluation des moteurs de recherche et méta-moteurs de recherche par rapport aux besoins des utilisateurs.

Après avoir sélectionné certains moteurs et méta-moteurs, il nous faut les évaluer afin de ne présenter aux utilisateurs qu'un petit nombre d'outils. Ceci leur permettra aux usagers d'exploiter toutes les fonctionnalités et de ne pas de perdre dans la multitude de produits disponibles sur le marché.

L'équation de recherche est assez spécifique et correspond à l'attente d'un chercheur du centre. Comme nous avons déjà travaillé sur cette question, il m'était plus facile d'évaluer la pertinence des résultats.

La même équation sera posée à chaque moteur suivant sa syntaxe spécifique. C'est pourquoi quand les opérateurs ne sont pas supportés par les moteurs, la question est divisée en trois.

(1) *Moteurs de type index.*

- **AltaVista** (recherche avancée): neurotrophin AND ((“ngf receptor” OR trk OR p75))
- **Excite** : neurotrophin AND ( ngf AND receptor OR trk OR p75)
- **HotBot** : neurotrophin AND “ngf receptor”  
neurotrophin AND trk  
neurotrophin AND p75
- **Infoseek** : neurotrophin AND “ngf receptor”  
neurotrophin AND trk  
neurotrophin AND p75
- **Lycos** : neurotrophin AND “ngf receptor ”  
neurotrophin AND trk  
neurotrophin AND p75
- **Magellan** : neurotrophin AND ngf AND receptor  
neurotrophin AND trk  
neurotrophin AND p75

(2) *Les moteurs de type répertoire.*

- Une première sélection a été faite à partir d’une liste déjà établie (Andrieu, 1997 ; Revelli, 1998). Nous avons ensuite regardé les parties “Science et Santé” afin d’évaluer les sites présentés. Il faut noter qu’aucun des outils cités ci-après n’est basé sur le même système de classification . Ceci permet d’avoir différentes vues sur le sujet et permet de mieux préparer une recherche sur un outil de type index.

(3) *Les méta moteurs.*

- **Metacrawler** : neurotrophin AND ngf AND receptor  
neurotrophin AND trk  
neurotrophin AND p75
- **Profusion** : neurotrophin AND (( ngf AND receptor ) OR trk OR p75)

Options lors de l'interrogation :

- Search mode : boolean
- Search in : Web
- Summary option : yes
- Check links : 10
- Moteurs : choose all

- **DogPile** : neurotrophin AND (“ngf receptor” OR trk OR p75)

Remarques :

- Attribue “un poids” à chaque terme.
- La présentation des résultats n'est pas satisfaisante : les résultats sont donnés pour chaque moteur de recherche, donc pas de dédoublement, et tous les moteurs ne sont pas sur la même page.

(4) *Résultats des recherches.*

(a) *Les moteurs de recherche de type index.*

Moteur	AltaVista	Excite	HotBot	Infoseek	Lycos	Magellan
Rapidité	Très bonne	Très bonne	Très bonne	Très bonne	Très bonne	Très bonne
Nombre de réponses	501	242	226- 404 - 160	2 255 495	11 - 36 -0	1717- 189 168
Nombre de réponses après affinage	120	-	-	-	-	-
Page de résultats (nombre de résultats par page)	10	10	10	10	10	10
Page de résultats (présentation des adresses)	* classement * URL * titre * résumé * date de dernière visite du fichier * taille du fichier * langue du document	* taux de pertinence * URL * titre * résumé * option « More like this »	* classement * URL * titre * résumé (bon) * taille du fichier	* pertinence * URL * titre * résumé * date de dernière visite du fichier * taille du fichier	* classement * URL * titre * résumé * taille du fichier * option « More like this »	* classement * URL * résumé * ville
Doublons (/30)	0	0	1- 1 -5	2- 0 - 1	0	0
Pourcentage de doublons (%)	0 (après affinage 8,33%)	0	0,88	0	0	0
Nombre de références non pertinentes (/30)	0	0	0	1	0	5 - 20 - 16
Date des adresses	les premières adresses datent de 1996, 1997 et 1993 !	?	?	?	?	?

Moteur	AltaVista	Excite	HotBot	Infoseek	Lycos	Magellan
Types d'adresses récupérées	<ul style="list-style-type: none"> <li>* congrès, séminaires, conférences</li> <li>* page personnelles</li> <li>* produits</li> <li>* articles</li> <li>* laboratoires</li> <li>* protocoles</li> <li>* bibliographies</li> </ul>	<ul style="list-style-type: none"> <li>* données cristallographiques</li> <li>* articles</li> <li>* base de données</li> <li>* pages personnelles</li> </ul>	<ul style="list-style-type: none"> <li>* brevets</li> <li>* entreprises</li> <li>* laboratoires</li> <li>* articles</li> <li>* cours</li> <li>* forums online</li> <li>* pages personnelles</li> </ul>	<ul style="list-style-type: none"> <li>* conférences</li> <li>* pages personnelles</li> <li>* articles</li> <li>* texte de loi</li> <li>* catalogue</li> <li>* page ressource</li> <li>* base de données</li> <li>* données cristallographiques</li> <li>* posters</li> </ul>	<ul style="list-style-type: none"> <li>* produits</li> <li>* articles</li> <li>* pages personnelles</li> <li>* articles de Medline</li> </ul>	<ul style="list-style-type: none"> <li>* symposium</li> <li>* données cristallographiques</li> <li>* articles</li> <li>* cours</li> </ul>

## (b) Les méta-moteurs.

Moteur	Metacrawler	Profusion	DogPile
Rapidité	Très bonne	Très bonne	Très bonne
Nombre de réponses	25	66	?
Nombre de réponses après affinage	-	-	-
Page de résultats (nombre de résultats par page)	10	en une page	moteur par moteur
Page de résultats (présentation des adresses)	<ul style="list-style-type: none"> <li>* pertinence</li> <li>* URL</li> <li>* résumé selon le moteur qui a trouvé l'adresse</li> <li>* date de dernière visite du fichier</li> <li>* nom du moteur</li> </ul>	<ul style="list-style-type: none"> <li>* taux de pertinence</li> <li>* URL</li> <li>* titre</li> <li>* résumé</li> </ul>	<ul style="list-style-type: none"> <li>* classement</li> <li>* URL</li> <li>* titre</li> <li>* résumé</li> </ul>
Doublons (/30)	0	1	-
Pourcentage de doublons (%)	0	0	-
Nombre de références non pertinentes (/30)	0	1	-
Date des adresses	?	?	-
Types d'adresses récupérées	<ul style="list-style-type: none"> <li>* congrès</li> <li>* données cristallographiques</li> <li>* programmes de recherche</li> <li>* articles</li> <li>* laboratoires</li> </ul>	<ul style="list-style-type: none"> <li>* données cristallographiques</li> <li>* laboratoires</li> <li>* symposium, congrès</li> <li>* projet de recherche</li> <li>* articles</li> <li>* base de données</li> <li>* pages personnelles</li> </ul>	

(5) *Exploitation des résultats.*

(a) *Les moteurs de type index.*

- **AltaVista** : les résultats obtenus sont intéressants de part leur diversité. Par contre, au niveau de la fraîcheur des sites récupérés, ce n'est pas au point ! Il y a peu d'adresses de 1998. D'autre part, la fonction d'affinage est utile mais il y a beaucoup de doublons dans la liste après cette opération.
- **Excite** : les types d'adresses sont peu variés.
- **HotBot** : ce ne sont pas les mêmes adresses qu'avec AltaVista (sauf 2 identiques). En séparant les équations de recherche, nous n'obtenons pas les mêmes adresses pour les trois cas, ce qui est intéressant. HotBot et AltaVista ne doivent pas avoir le même index.
- **Infoseek** : il y a 63 nouvelles adresses par rapport aux deux interrogations précédentes. Par contre, le chiffre annoncé de plus de 2 millions de pages trouvées pour cette question est assez bizarre.
- **Lycos** : le nombre de pages est peu important et la diversité des types de fichiers est faible.
- **Magellan** : cet outil ne s'avère pas approprié aux recherches dans son index. D'abord par le peu de types de fichiers qu'il offre et par le nombre assez important de sites qui ne correspondent pas à notre recherche.

(b) *Les moteurs de type répertoire.*

Les résultats obtenus sont les suivants :

- Nomade : sélection de sites francophones.
- Yahoo : incontournable, donne de nombreux cours, journaux, noms d'organisations .
- Magellan : ce moteur ne répertorie pas beaucoup de sites mais présente de très bons sites ressources.

- ClearingHouse : ce moteur présente une sélection très restreinte de sites et il est très long à charger. C'est pourquoi nous ne sélectionnerons pas ce moteur mais nous nous en inspirerons pour faire la partie domaine du site d'accueil.

Galaxy : idem que pour ClearingHouse.

(c) *Les méta-moteurs.*

Metacrawler donne peu de sites comparé à Profusion.

DogPile n'est pas facile d'utilisation : comme les résultats sont donnés par moteur de recherche, les doublons ne sont pas enlevés et il faut parcourir de nombreuses pages pour voir les derniers résultats. L'autre inconvénient est que tous les moteurs de recherche appelés n'ont pas répondu !

Enfin, il n'y pas la date de dernière visite des fichiers dans Profusion.

(6) *Conclusions.*

Par rapport aux résultats de l'évaluation de chaque moteur de recherche et de chaque méta-moteur, nous avons décidé de choisir :

- comme moteur de recherche de type index: **AltaVista, HotBot et Infoseek.**
- comme moteur de recherche de type répertoire : **Nomade, Yahoo (américain), Magellan.**
- comme méta-moteur : **Profusion.**

- e) Description détaillée des caractéristiques des moteurs de recherche sélectionnés.

(1) *Les moteurs de recherche de type index.*

Les caractéristiques des moteurs sélectionnés proviennent des sites des moteurs ainsi que de certains guides en ligne (Barlow, 1997 ; Charron, 1997 ; Sullivan, 1998).

(a) *Description du moteur de recherche.*

- URL du moteur.
- Taille de l'index.
- Fréquence de mise à jour : montre l'âge des index.

(b) *Collecte des documents.*

- Nom du robot pour la collecte.
- Fichiers collectés.
- Couverture géographique.
- Nombre de pages indexées par jour : montre combien de pages le robot peut visiter par jour. Plus il visite pages, plus l'index est actualisé, donc plus l'information peut être estimée " fraîche".
- Pages soumises : délais entre le moment où l'on a soumis la page à indexer au moteur et le moment où elle est réellement indexée.
- Pages non soumises : délais entre le moment où le robot a visité et indexé une page et le moment où il a indexé le reste du site ou l'échantillon de pages web.
- Profondeur : en relation avec les pages non soumises. Présente le nombre de pages non soumises que le robot va indexer:
  - \* illimité : le robot va indexer un maximum de pages.
  - \* échantillon : le robot n'indexe qu'un échantillon de pages.
- Cadres : indique si le robot peut suivre les liens présents dans le cadre.
- Sites protégés par mot de passe : indique si le robot peut rentrer dans des sites protégés par mot de passe.
- Pénalités pour "spam" : les moteurs de recherche analysent la densité de mots (nombre d'occurrence du mot demandé divisé par le nombre de mots total de la page). Si ce rapport est trop élevé alors la page sera rejetée pour spam.

Qu'est ce que le "spam" ? C'est l'altération ou la création d'un document avec l'intention de tromper par exemple un catalogue électronique ou des moteurs de recherche. Concrètement dans le cas des moteurs de recherche, les "spammeurs"

utilisent la répétition de mots-clés, des duplications de pages avec des URLs différentes, le texte de la même couleur que le fond de la page, ...

(c) *Indexation des documents.*

- Méthode d'indexation : manuelle ou automatique.
- Nom du moteur d'indexation : si indexation automatique.
- Données indexées.
- Méthode d'élimination des mots vides.

(d) *Recherche de documents.*

- Nom du moteur de recherche.
- Opérateurs disponibles.

Le système +/- : mettre le signe + devant un terme, signifie que l'on veut que le mot soit dans le document. Au contraire, un signe - signifie que l'on ne veut pas que le mot soit dans le document.

- Sensible aux majuscules :

Le "oui" indique que quand le mot est en majuscule, le moteur va le chercher tel quel. S'il est en minuscule, le moteur cherchera le mot en minuscule et en majuscule.

- Troncature : seul AltaVista permet la troncature manuelle. Pour cela , utiliser l'étoile pour tronquer trois caractères maximum.

- Accentuation :

Le "oui" indique que si le caractère est accentué, le moteur ne cherchera que les caractères accentués, sinon il cherchera les caractères accentués et non accentués.

- Recherche de phrase : pour rechercher par phrase, utiliser les mots entre guillemets.
- Recherche dans les champs : liste de tous les champs, ainsi que de la syntaxe dans lesquels il est possible de faire une recherche spécifique.
- Méthode de tri des résultats.
- Tri des résultats.
- Amélioration de la recherche.

(e) *Présentation des résultats.*

- Titre : explique d'où est extrait le titre du document.
- Description : explique comment est établit la description du document.
- Date.
- Résultats / page : traduit le nombre de résultats par page.
- Options pour la présentation : décrit les possibilités offertes par le moteur de recherche pour la présentation des documents.

	<b>ALTAVISTA</b>	<b>HOTBOT</b>	<b>INFOSEEK</b>
<b>DESCRIPTION</b>			
URL	www.altavista.digital.com	www.hotbot.com	www.infoseek.com
Taille de l'index (pages)	140 000 000	110 000 000	30 000 000
Fréquence de mise à jour	1 jour à 1 mois	1 jour à 2 semaines	1 jour à 2 mois
<b>COLLECTE</b>			
Robot	Scouter	Inktomi	Side Winder
Fichiers collectés	WWW, usenet, news	WWW, usenet, news	WWW, gopher, ftp, usenet, PrNewswire, CSA Biomedical database
Couverture géographique	mondiale	mondiale	mondiale
Nombre de pages visitées par jour	10 000 000	10 000 000	-
Pages soumises	1 jour	1 jour à 2 jours	1 jour
Pages non soumises	1 jour à 1 mois	2 semaines	1 à 2 mois
Profondeur	illimité	illimité	échantillon
Cadres supportés	non	non	oui
Pages obsolètes	enlève les pages et resoumet les nouvelles	créer un fichier qui liste les pages à enlever puis détruit les pages	deux systèmes
Sites protégés par mot de passe	non	non	oui
Pénalité pour spam	oui	oui	oui

	<b>ALTAVISTA</b>	<b>HOTBOT</b>	<b>INFOSEEK</b>
<b><i>INDEXATION</i></b>			
Méthode d'indexation	automatique, soumission d'URL	automatique, soumission d'URL	automatique, soumission d'URL
Moteur d'indexation	Indexer (Digital)	?	?
Données indexées	texte intégral	texte intégral	texte intégral
Méthode d'élimination des mots vides	fréquence des mots	inconnue	les mots les plus courants de l'anglais
<b><i>RECHERCHE</i></b>			
Moteur de recherche	Indexer (Digital)	Inktomi	?
Opérateurs	and, or, not, and not +/-, ( )	and, or, not, and not +/-, ( )	+/-, virgule (séparation des termes)
Opérateur par défaut	or	and	or
Opérateur de proximité	near (10 mots)	-	(pipe) = near
Troncature	manuelle :* à droite : 3 caractères interne	non	non
Sensible aux majuscules	oui (détail)	oui	oui (détail)
Gestion de l'accentuation	oui	non	non
Phrase	“ ”	“ ”	“ ”
Recherche dans les champs	date, langue title : domain :, host : url : anchor : applet : link : object : (marque) text : MULTIMEDIA : image :, sound :, java : USENET : from :, subject :, summary :, newsgroup : keywords :	title : link : date = after :, before :, domain : linkdomain : linkext : scriptlanguage : newsgroup : feature : (video, applet, image, audio, acrobat, vrml, ...) MULTIMEDIA : image :, sound :, java :	title : url : site : link :

	<b>ALTAVISTA</b>	<b>HOTBOT</b>	<b>INFOSEEK</b>
Quelques critères de tri des résultats	selon le nombre de termes de la requête apparaissent dans le document, où et leur proximité entre eux, ...	poids plus élevé des mots qui sont dans les titres, fréquence des mots dans le texte (plus élevé si le texte est court), ...	à partir de la fréquence, ...
Amélioration de la recherche	oui (fonction REFINE)	non	non
<b>PRESENTATION DES RESULTATS</b>			
Titre	titre de la page ou “ No Title ”	titre de la page ou URL	titre page ou première ligne de la page
Description	metatags et les 1eres lignes de la page	metatags et les 1eres lignes de la page	metatags ou 200 caractères après la balise <body>
Date	celle de la dernière visite du fichier	aucune	celle de la dernière visite du fichier
Résultats / page	10	10	10
Option pour la présentation	standard compact texte seul	complet bref (1 ligne) titres seuls	résumés titres seuls

(2) *Les moteurs de recherche de type index répertoire.*

Les trois moteurs de recherche de ce type que nous avons choisis sont intéressants car leur système de classification est différent.

- **Magellan** : comprend une rubrique “Santé - Médecine” et “Science”. Les découpages à l’intérieur de ces deux rubriques sont faits par spécialités pour lesquels sont répertoriés des sites ressources. Les sites sont en langue anglaise.
- **Nomade** : a une rubrique “Science” découpée selon les domaines scientifiques, les conférences, les technologies générales ou spécialisées, des informations de références, etc. Ce moteur ne répertorie que des sites en langue française avec une mention grand public ou professionnel.

- **Yahoo** : possède une rubrique “Santé” et “Science”. Le découpage est ensuite très fin et par spécialités. A l’intérieur de ces spécialités, le découpage est encore assez fin, jusqu’à donner, par exemple, des liens vers sites sur des médicaments très précis !

Quand faut-il se servir d’un type particulier? Si l’usager cherche des sites francophones, il peut utiliser Nomade. S’il cherche des sites ressources en science ou en santé, il peut aller consulter Magellan. Enfin, s’il cherche des sites assez spécifiques, il peut essayer de les trouver sur Yahoo.

(3) *Le méta-moteur de recherche.*

	<b>PROFUSION</b>
<b>URL</b>	http ://profusion.ittc.ukans.edu
<b>Moteurs de recherche utilisés</b>	AltaVista Excite HotBot Lycos Webcrawler Infoseek Magellan OpenText Yahoo
<b>Opérateurs disponibles (en majuscules)</b>	AND, OR, NOT, NEAR
<b>Comportement de chaque moteur par rapport aux opérateurs booléens</b>	<u>Supporte les booléens</u> : AltaVista Excite (l’opérateur NEAR est converti en AND) HotBot Lycos Webcrawler (l’opérateur NEAR est converti en AND) <u>Ne supporte pas les booléens</u> : Infoseek Magellan OpenText Yahoo
<b>Recherche de phrase</b>	“ ”
<b>Parenthésage</b>	oui
<b>Nombre de résultats par page</b>	tous les résultats sur une même page

	<b>PROFUSION</b>
<b>Présentation des résultats</b>	taux de pertinence, titre de la page, URL et résumé
<b>Options</b>	<ul style="list-style-type: none"> <li>• possibilité de choisir :  “ les 3 meilleurs moteurs ”  “ les 3 plus rapides ”  “ tous les moteurs ”  choix manuel</li> <li>• résultats avec ou sans résumé</li> <li>• vérifier les liens contenus dans les pages rapatriées (nombre de liens à spécifier)</li> <li>• possibilité d'enregistrer sa requête et de recevoir les nouveaux sites automatiquement</li> </ul>

(4) *Conclusion quant à la démarche de recherche à adopter.*

Pour conduire une bonne recherche sur Internet, je conseillerai de commencer avec un outil de type répertoire. La classification des adresses ainsi qu'une possibilité de chercher dans la rubrique et leur "index" est assez pratique. Ensuite, la question peut être posée sur Profusion, le méta-moteur. Enfin, si l'utilisateur veut une recherche plus ou moins exhaustive, une recherche sur un sujet pointu ou qu'il n'a pas trouvé des sites intéressants avec les deux outils précédents, il peut lancer la requête sur les trois moteurs de recherche de type index.

### 3. Recherche par type d'informations.

#### a) Objectif.

Cette partie de la méthodologie doit permettre à l'utilisateur de trouver un type d'information précis.

Pour cela, une structure particulière sera adoptée :

- une rubrique "**quelques sites**" qui présente des liens vers des sites ressources ou des exemples de ce que l'on peut trouver sur Internet.

- une rubrique “**moteurs spécialisés**” regroupant des liens vers des moteurs spécialisés, c’est à dire dont le contenu des documents indexés correspond au type d’information recherché, soit des liens vers des sites traitant du type d’information et sur lequel il existe une fonction de recherche.
- une rubrique “**exemple d’interrogation**” comprenant un lien dynamique vers le moteur AltaVista avec une équation correspondante.

b) Les différents types d’informations.

(1) *Sélection.*

Je me suis appuyée sur les résultats de l’enquête et inspirée des sites Sapristi (Joly, Bion, Mermet et *al.*, 1995) et GIRI 2 (Crepuq, 1997) afin d’établir une liste des types d’informations à traiter.

Je n’ai volontairement pas mis les ouvrages ni les pré-publications. Pour les ouvrages, il en existe peu en texte intégral sur Internet et ce ne serait pas très utile. Concernant les pré-publications, malgré les possibilités qu’offre Internet dans ce domaine, j’ai souhaité réduire le risque de désinformation possible pour les usagers.

(2) *Les types d’informations.*

- Administration : regroupe des sites administratifs, gouvernementaux, juridiques, des sites d’universités et de laboratoires de recherche.
- Adresse électronique ou postale.
- Banque de données : celles accessibles sur Internet.
- Brevet .
- Conférence, congrès, séminaire : des annonces, des comptes-rendus.
- Cours.
- Dictionnaire.
- Données de référence : comprend des sites regroupant des constantes, des protocoles, ...

- Faq ou “ Frequently Asked Questions ”.
- Ftp ou “ File Transfer Protocol ”.
- Image.
- Liste et forum de discussion.
- Logiciel.
- Norme et des organismes de normalisation.
- Organisation : permet de chercher des organisations, des institutions, des sites de sociétés savantes.
- Page de liens ou sites ressources.
- Périodique : ceux présents sur Internet.
- Rapport : rapports de sociétés, d’activité.
- Société, produit et fournisseur.
- Thèse : recherche de référence bibliographique ou de thèse en texte intégral.

c) Sélection des liens et des moteurs spécialisés.

La sélection s’est faite grâce aux sites Saprستي et GIRI 2, à des ouvrages (Andrieu, 1997 ; Revelli, 1998), aux listes de diffusion (Adbs- info ; Biblio-fr ; Bibliosante ; Metanews ; Moteurs de recherche ; Veille ), lors de la navigation sur le Web et enfin par l’interrogation des moteurs de recherche.

Nous présenterons succinctement les moteurs spécialisés et leur syntaxe d’interrogation.

d) L’équation de recherche et AltaVista.

C’est AltaVista qui a été choisi pour ses possibilités lors des interrogations et la taille de son index (140 millions de page). L’exemple de l’équation de recherche a été inspiré du site Saprستي, mais j’y ai apporté des modifications tant au niveau des termes que de la possibilité technique pour l’interroger.

En effet, sur le site Saprستي, l’équation de recherche est pré-enregistrée sur le site miroir d’AltaVista en Norvège. Pour des raisons de taille d’index (celui du site miroir est largement inférieur à celui situé en Californie) et des raisons de maintenance, j’ai choisi

de développer un système qui permet de copier l'équation dans la zone de recherche avancée d'AltaVista (site en Californie), de modifier la partie "sujet/mot-clé" par les termes de l'utilisateur et d'envoyer la requête. Concernant les termes de la recherche, j'ai cherché à être un peu plus exhaustive en utilisant, entre autre, la fonction d'affinage d'Alta Vista (fonction Refine) qui permet de voir les termes associés à notre requête.

Un exemple d'interrogation :

- Sélectionnez l'équation de recherche *correspondant au moteur de recherche choisi, puis copiez la avec le bouton droit de la souris.*
- Cliquez sur l'image *du moteur de recherche.*
- *Dans la zone de recherche "Boolean Expression", collez l'équation avec le bouton droit de la souris.*
- Modifiez "sujet/mot-clé" *par votre ou vos mots.*

*Conseil: essayez d'élargir votre recherche, en ne cherchant pas vos mots- clés dans le titre, si les résultats ne vous conviennent pas.*

- Lancez la recherche.

Moteur de recherche (lien dynamique)	Equation de recherche
	<p>title:sujet/mot-clé and (title:conference* or title:proceeding* or title:semina* or title:workshop* or title:sympos* or title:meeting* or title:congres* or title:colloqu*)</p>

J'avais aussi prévu de développer ce système avec les moteurs de recherche HotBot et Infoseek. Concernant HotBot, la difficulté réside dans le fait que, d'une part, il ne supporte pas la troncature et d'autre part, problème assez ennuyeux, la zone de recherche est limitée en nombre de caractères. Ceci empêche le développement d'une longue équation de recherche. Quant à Infoseek, le système d'opérateurs +/- est limité

par rapport aux opérateurs booléens et son index est moins volumineux. Enfin, je pense qu'il ne faut pas trop multiplier les accès aux moteurs au risque de dérouter les utilisateurs.

## **C. LA SELECTION DE SITES PAR DOMAINES.**

### **1. Objectif.**

Cette partie du site d'accueil doit permettre à l'utilisateur d'avoir des points d'entrée vers des sites connus et importants en biotechnologie ou des sites très souvent utilisés.

### **2. Sélection des domaines.**

Cette sélection des sites s'est faite à partir des différentes activités du centre de recherche et des résultats de l'enquête.

Les domaines sélectionnés sont:

- Biochimie
- Biologie Moléculaire
- Documentation
- Législation, Qualité, Réglementation
- Médecine, Santé
- Microbiologie
- Neurobiologie
- Protocoles, Techniques de Laboratoires
- Sécurité
- Substances Naturelles

### 3. Sélection des sites.

Elle s'est faite grâce à des adresses obtenues sur les listes de discussion (Abds-info ; Biblio-fr ; Bibliosanté), grâce à des ouvrages (Crepuq, 1997 ; Rosenfeld, Janes et Vander Holk, 1995 ; Urfist, 1997 ; Wayne, 1997) et des articles (Lomme, 1998 ; Thirion et Darmon, 1998). Je me suis également servie de la partie recherche d'information et des équations de recherche.

### 4. Liens vers la rubrique "Web Alert " des périodiques.

J'ai découvert grâce à un utilisateur une rubrique intitulée "Web Alert " dans une série de périodiques intitulée "Current Opinion in ...". C'est une sélection de sites Internet par domaines et par spécialités, qui paraît à chaque numéro.

J'ai donc dépouillé les versions papier et recherché les titres présents en texte intégral sur Internet. L'accès au texte intégral des articles est payant mais cette rubrique est disponible pour tout un chacun.

Après une discussion avec Mme Emerat, il nous a paru intéressant de fournir, par domaines, un accès à ces sélections.

Le lien est par défaut sur le dernier numéro du périodique, mais il est possible de rechercher dans les numéros antérieurs.

Les titres sélectionnés sont :

- Current Opinion in Biotechnology
- Current Opinion in Cell Biology
- Current Opinion in Chemical Biology
- Current Opinion in Genetics and Development
- Current Opinion in Immunology
- Current Opinion in Neurobiology
- Current Opinion in Neurology
- Current Opinion in Structural Biology

## **D. LA CHARTE GRAPHIQUE.**

### **1. Les règles générales de développement.**

La charte graphique suit les mêmes règles générales de développement que l'Intranet de Sanofi.

Ce qui suit présente uniquement les modifications apportées.

- Navigateur à utiliser : Netscape 4.04
- Taille de l'écran : le contrôle se fait sur un écran 15 pouces.

### **2. Présentation générale du site.**

Le site se compose de quatre parties : l'accueil, les journaux en texte intégral, la partie regroupant les domaines et la partie concernant la recherche d'informations comprenant la présentation des moteurs de recherche et l'aide à la recherche par type d'informations.

Je présenterai les règles de développement en séparant les différentes parties des pages (cadre et page principale).

### **3. L'accueil.**

- Le cadre :
  - \* Fond : couleur #A3D1D1
  - \* Applets Java :
    - code : fphover.class
    - code base : \_fpclass
    - largeur, hauteur : 170, 24
    - couleur : #A3D1D1
    - effet : lumière
    - police : Times New Roman, gras, 14 pt, couleur #FFFFFF

- Page principale :
  - \* Fond : couleur #FFFFFF
  - \* Titre : Comic Sans Ms, gras, 24 pt, couleur #FF0000
  - \* Tableau : Times New Roman, normal, 14 pt, couleur du fond #A3D1D1, couleur du texte et des liens #000000
  - \* Images :
    - bordure : 0
    - largeur, hauteur : 44, 75
    - cadre de destination : page entière

#### 4. Les journaux.

- Pas de cadre.
- Page principale :
  - \* Fond : couleur #FCFED8
  - \* Titre : Comic Sans Ms, gras, 18 pt, couleur #FF0000
  - \* Tableau : Times New Roman, normal, 14 pt, couleur du fond #FCFED8, couleur du texte et des liens #000000
  - \* Outil de navigation : flèche pointant vers la gauche.
    - Propriétés de l'image :
      - bordure : 0
      - chemin d'accès dans Front Page :  
images/arrow\_big\_left.gif
      - texte : Retour vers l'accueil
      - cadre de destination : page entière
  - \* Date de mise à jour : composant Webbot type Webstamp, format :  
jour/mois/année.

## 5. Les domaines.

- Le cadre : c'est le même que celui de la page d'accueil mis à part que le domaine consulté est mis en évidence par un changement de couleur de l'applet qui est #008080.
- Page principale :
  - \* Fond : couleur #FFFFFF
  - \* Titre : Comic Sans Ms, gras, 18 pt, couleur #40ACFD
  - \* Tableau : regroupe les liens vers les différentes rubriques de la page.  
bordure de taille 2, alignement : centré  
texte : Comic Sans Ms, normal, 14 pt, couleur du fond #FFFFFF, couleur du texte et des liens #008080.
  - \* Rubriques:
    - texte du signet: nom de la rubrique, Comic Sans Ms, gras, 14 pt, couleur du texte #FF0000
    - texte de la rubrique : Times New Roman, normal, 12 pts, couleur par défaut.
    - ligne horizontale pour séparer les rubriques.
  - \* Outil de navigation : flèche pointant vers le haut à chaque rubrique.  
Propriétés de l'image :
    - bordure : 0
    - chemin d'accès dans Front Page :  
images/arrow\_big\_up.gif
    - texte : Retour vers le haut
    - lien vers le titre par un signet
  - \* Date de mise à jour : composant Webbot type Webstamp, format : jour/mois/année.

## 6. La recherche d'informations.

- Pour toutes les pages :
  - \* Fond : couleur # FCFED8
  - \* Titre : Comic Sans Ms, gras, 18 pt, couleur # FF0000
  - \* Date de mise à jour : composant Webbot type Webstamp, format : jour/mois/année.
  
- Page d'orientation regroupant tous les types d'informations .
  - \* Tableau : composé avec Microsoft Image Composer, l'éditeur d'images. Les images sont en format jpeg, chacune pointant vers la page correspondante et le cadre de destination est une page entière.
  - \* Outil de navigation : flèche pointant vers la gauche.
    - Propriétés de l'image :
      - bordure : 0
      - chemin d'accès dans Front Page :  
images/arrow\_big\_left.gif
      - texte : Retour vers l'accueil
      - cadre de destination : page entière, lien vers la page d'accueil du site.
  
- Pages des différents types d'informations.
  - \* Titre : Comic Sans Ms, gras, 18 pt, couleur # FF0000.
  - \* Outils de navigation :
    - ⇒ flèche pointant vers la gauche.
      - Propriétés de l'image :
        - bordure : 0
        - composée sur Microsoft Image Composer, flèche blanche sur un fond bleu pour rappeler le tableau des types d'informations.
        - texte : Recherche par type d'informations.

- cadre de destination : page entière, lien vers la page du tableau.

⇒ flèche pointant vers le haut à chaque rubrique.

Propriétés de l'image :

- bordure : 0

- chemin d'accès dans Front Page :  
images/arrow\_big\_up.gif

- texte : Retour vers le haut de la page.

- lien vers le titre par un signet

\* Signets de départ présentant les trois rubriques "quelques sites", "les moteurs spécialisés" et "l'exemple d'interrogation".

Police des signets : Comic Sans Ms, gras, 14 pt, couleur # 3777A88.

Chaque signet a un lien hypertexte vers la rubrique correspondante.

\* Signets de chaque rubrique.

- Nom du signet : une des trois rubriques

- Police : Times New Roman, gras, 14 pts, couleur par défaut.

\* Texte : Times New Roman, normal, 12 pts, couleur par défaut.

\* Tableau de la rubrique "Exemple d'interrogation"» :

- bordure : 0

- police de l'équation : Times New Roman, gras, 12 pts, couleur #FF0000 sauf la partie "sujet/mot-clé" qui est en #0080FF.

### **III. REALISATION ET MISE EN OEUVRE .**

#### **A. LES MOYENS INFORMATIQUES.**

##### **1. Le logiciel FrontPage.**

J'ai utilisé le logiciel FrontPage 98 édité par Microsoft (Randall et Jones , 1997). Ce logiciel est composé de plusieurs éléments.

###### **a) L'explorateur.**

Il permet la création d'un nouveau site, l'importation de documents, la configuration et la gestion du site.

Il présente le site que l'on a créé sous différents affichages (les principaux):

- l'affichage Liens hypertextes : présente les liens hiérarchiques et les liens entre les pages.
- l'affichage Navigation : présente également la hiérarchie du site mais permet aussi de le réorganiser en vue de faciliter la navigation.
- l'affichage Dossiers : montre les dossiers comme dans l'Explorateur de Windows.
- l'affichage Etat des liens hypertextes : présente l'état des liens, ceux qui fonctionnent et ceux qu'il faut modifier.
- l'affichage Tâche : permet de voir la liste des tâches restantes avec la possibilité de les attribuer à certaines personnes.

###### **b) L'éditeur.**

Cet élément permet d'écrire les pages en langage HTML (Hypertext Markup Language) sans en avoir une très bonne connaissance. Il permet aussi de gérer les tableaux, les cadres, les liens vers les bases de données, les scripts, l'HTML dynamique, l'insertion de contrôles ActiveX et d'applets Java , et tout ce qui est nécessaire à la présentation visuelle.

C'est un logiciel WYSIWYG (what you see is what you get) mais j'ai eu des surprises !

c) Le serveur Web personnel.

Permet de gérer le site Web et de le rendre accessible aux utilisateurs.

d) Microsoft Image Composer.

Cet élément permet de créer des images, de les modifier même si elles viennent de l'extérieur. Les images en sortie sont en format JPEG, GIF ou MIC (format propre à Microsoft : Microsoft Image Composer).

La navigateur utilisé pour visualiser le site est Netscape Navigator 4.04.

## **B. LA CREATION DU SITE.**

### **1. L'architecture du site.**

Le site est développé en local sur mon poste de travail.

<http://w00136docnet/sitelabege>

- `_private` : sert à stocker les fichiers de résultats si l'on insère des scripts.
- Domaines : dossier contenant toutes les pages html ainsi que les cadres correspondant à la partie "Quelques sites".
- Images : dossier propre à Microsoft où sont presque toutes les images nécessaire au site. Attention, si les images ne sont pas rangées dans ce dossier, on observe des choses bizarres.
- Méthodologie : dossier regroupant toutes les pages html, les cadres de la méthodologie et l'image "Retour" créée avec Image Composer.
- Les pages de l'accueil général
- La liste des journaux en texte intégral.

## **2. Les pages html.**

Elles ont été conçues suivant le cahier des charges avec l'éditeur de Front Page et visualisées sous Netscape Navigator 4.04.

## **3. Des exemples de pages.**

Dans les pages qui suivent, je vous présente des exemples de chaque partie du site :

- L'accueil.
- Biochimie : présente un exemple de page de la partie " Domaines".
- Les moteurs de type index : un exemple de présentation de moteurs de recherche.
- Le tableau de la partie recherche d'informations par type.
- Conférences : un exemple de recherche par type d'information.

**Notre sélection de sites:**

[Biochimie](#)

[Biologie Moléculaire](#)

[Documentation](#)

[Médicament - Sanofi](#)

[Microbiologie](#)

[Neurobiologie](#)

[Pathologie - Vaccins](#)

[Pharmacie - Réglementation](#)

[Pharmacologie](#)

[Produits Naturelles](#)

[Produits de Sanofi](#)



[Journaux en texte intégral](#)

[Comment rechercher une information?](#)

[En savoir plus sur les moteurs de recherche.](#)

# SID LABEGE

- bienvenue sur internet -

Accès direct aux moteurs de recherche:



[Profusion](#)



**MAGELLAN**  
INTERNET GUIDE

[Index](#)

[Sanofi Net](#)

[Les directives de sécurité pour Internet](#)

Notre sélection de sites:

Biochimie

http://www.via.ecp.fr

## BIOCHIMIE

<u>Associations</u>	<u>Journaux</u>
<u>Banques de données</u>	<u>Nomenclature</u>
<u>Conférences</u>	<u>Ressources</u>

### Associations

- [American Society for Biochemistry and Molecular Biology.](#)
- [American Chemical Society.](#)
- [Royal Society of Chemistry](#)

### Banques de données





## PRESENTATION DES MOTEURS DE TYPE INDEX

Moteur	AltaVista	HotBot	Infoseek
Lien dynamique vers le moteur de recherche			
Taille de l'index (nombre de pages)	140 000 000	110 000 000	30 000 000
Mise à jour	1 mois au maximum	2 semaines au maximum	2 mois au maximum
Serveurs collectés	WWW, usenet, news	WWW, usenet, news	WWW, gopher, ftp, usenet, PRNewsWire, CSA Biomedical Database
Opérateurs			
Booléens	and, or, not, and not	and, or, not, and not	aucun
Système +/-	+ : mot doit être dans le document - : mot ne doit pas être dans le	+ : mot doit être dans le document - : mot ne doit pas être dans le	+ : mot doit être dans le document - : mot ne doit pas être dans le

<i>Parenthèses</i>	oui	oui	non
<i>Troncature</i>	*	non	automatique
<i>Phrase</i>	" "	" "	" "
<i>Sensible aux majuscule</i>	oui	oui	oui
<i>Sensible aux accents</i>	non	oui	oui
<i>Recherche dans les champs spécifiques</i>	<ul style="list-style-type: none"> <li>title:</li> <li>domain:</li> <li>host:</li> <li>url:</li> <li>anchor:</li> <li>applet:</li> <li>link:</li> <li>object (marque):</li> <li>text:</li> </ul>	<ul style="list-style-type: none"> <li><u>Multimédia:</u></li> <li>image:</li> <li>sound:</li> <li>java:</li> <li><u>Usenet:</u></li> <li>from:</li> <li>subject:</li> <li>summary:</li> <li>newsgroup:</li> <li>keyword:</li> </ul>	<ul style="list-style-type: none"> <li>title:</li> <li>domain:</li> <li>link:</li> <li>date=after/before:</li> <li>link domain:</li> <li>linkext:</li> <li>scriptlanguage:</li> <li>newsgroup:</li> <li>feature(video, applet, image, audio, acrobat, vml):</li> <li><u>Multimédia:</u></li> <li>image:</li> <li>sound:</li> <li>java:</li> <li>title:</li> <li>url:</li> <li>site:</li> <li>link:</li> </ul>

Dernière mise à jour: 31/08/98



## RECHERCHE PAR TYPE D'INFORMATIONS

administration	adresse	banque de données	brevet
conférence	cours	dictionnaire lexique	données de référence
faq	ftp	image	liste et forum de discussion
logiciel	norme	organisation	page de liens
périodique	rapport	société produit	thèse

Index - Notes aux utilisateurs



## CONFERENCE, SEMINAIRE, CONGRES

Vous cherchez l'annonce d'un congrès, d'une conférence ou encore les comptes-rendus de ces événements?

### Quelques sites - Les moteurs spécialisés - Exemple d'interrogation

#### Quelques sites



- Annonces de conférences par l'Institut Pasteur: liste proposée par l'Institut Pasteur.
- MedWeb: conférences par spécialités
- The Calendar of Upcoming Technical Conferences
- Liste de conférences en biologie (université de Waterloo, Canada)
- Les conférences ayant lieu en France
- Liste de conférences en biotechnologie
- Liste proposée par Cell and Molecular Biology online

## Les moteurs spécialisés ▲

	<i>Ce que vous y trouverez.</i>	<i>Sources d'informations.</i>	<i>Comment interroger?</i>
<u>Medical Conference</u>	7 000 annonces de conférences	Recherche par mot-clés, par titre de conférence, par date ou par lieu.	
<u>INIST</u>	comptes rendus	INIST	recherche par mots du titre, numéro d'identifiant, le numéro de côte de l'INIST, nom d'auteur, et par langue.  opérateurs: <AND>, <IN>, <NEAR>, <NOT>, <OR>, ... troncature: * pour n caractères, ? pour 1 caractère dans le mot aide: <b>en ligne</b>

## Exemple d'interrogation



- Sélectionnez l'équation de recherche correspondant au moteur de recherche choisi, puis copiez la avec le bouton droit de la souris.
- Cliquez sur l'image du moteur de recherche.
- Dans la zone de recherche "Boolean Expression", collez l'équation avec le bouton droit de la souris.
- Modifiez "sujet/mot-clé" par votre ou vos mots.  
*Conseil: essayez d'élargir votre recherche, en ne cherchant pas vos mots-clés dans le titre, si les résultats ne vous conviennent pas.*
- Lancez la recherche.

Moteur de recherche	Equation de recherche
	title:sujet/mot-clé and (title:conference* or title:proceeding* or title:semina* or title:workshop* or title:sympos* or title:meeting* or title:congres* or title:colloqu*)

### **C. LA CONFIGURATION DU SITE.**

La configuration du site est celle par défaut du logiciel.

Concernant les différentes autorisations :

- administration et modification : c'est le personnel du service d'Information Documentaire et plus particulièrement Mme Emerat, qui sera administrateur et aura la possibilité de modifier le contenu du site.
- l'exploration : les utilisateurs qui ont un compte déclaré pour Internet auront la possibilité de naviguer sur le site, en lecture seule.

### **D. LA GESTION DU SITE.**

Les documents présents sur Internet se modifient vite, les liens deviennent aussi rapidement obsolètes et les usagers se lassent. Ceci implique la nécessité d'une bonne maintenance et d'une mise à jour constante des informations disponibles sur le site. Toute cette partie sera prise en main par le service d'Information Documentaire.

#### **1. La maintenance.**

##### **a) Les lien hypertextes.**

La maintenance des liens passe par deux étapes.

- La commande "Recalculer des liens" permet, entres autres, une fonction très utile : l'actualisation des affichages de l'explorateur, important si on a modifié des liens, ajouté ou modifié des pages.
- La commande "Vérification des liens hypertextes" de l'explorateur Front Page. Elle permet de vérifier la validité des liens internes au site. A partir de la boîte de dialogue, il est possible de modifier la page ou de la placer dans la liste des tâches à faire. Ensuite, si on clique sur « Vérifier », le logiciel FrontPage va contrôler l'état des liens qui pointent vers des adresses externes.

Ces deux étapes devraient être réalisées au minimum une fois par mois, l'une après l'autre.

b) Le serveur, les ordinateurs.

Cette maintenance sera effectuée par le service informatique.

## **2. La mise à jour.**

La mise à jour va permettre d'éviter que les usagers ne se lassent du site et qu'ils y trouvent des nouvelles adresses, des nouvelles rubriques.

Cette mise à jour au niveau du contenu passe par une veille constante sur les listes de discussion appropriées, par une navigation et des recherches d'informations assez fréquentes. Au niveau de la forme, il serait utile de pouvoir changer de couleurs, d'ajouter des rubriques, des nouveautés, afin d'éveiller la curiosité de l'utilisateur.

#### IV. CONCLUSION.

Grâce à ce projet, j'ai pu mettre à profit ma double compétence documentaire et scientifique. Mon contact permanent avec les utilisateurs m'a permis de jouer le rôle d'un vrai prestataire de services. Cette relation a été très enrichissante tant au niveau professionnel que personnel.

Le logiciel FrontPage n'a presque plus de secrets pour moi mais il n'est pas très facile à manier. De plus, comme il intègre toutes les fonctions de création, maintien et gestion d'un site, il se rend quasiment indispensable !

Le site va prochainement être présenté et proposé aux utilisateurs d'Internet. Il serait souhaitable de refaire une enquête dans les mois à venir afin d'avoir un retour quant à l'aide apportée, pour cibler les objectifs du site et proposer de nouveaux services.

La mise en place de profils personnalisés sur Internet est un service à développer rapidement. Rappelons nous que 80 pour cent des utilisateurs souhaitaient y avoir accès ! Suite à cette étape, le SID pourrait développer un système de push des résultats des profils *via* l'intranet du centre.

Ce travail, son suivi ainsi que les futurs développements à mener, montre bien le rôle central que nous avons à jouer. Nos connaissances en matière de sources d'informations disponibles, d'outils de recherche (moteurs, agents et logiciels de veille) et d'outils de mise à disposition de l'information, deviennent indispensables pour aider les internautes.

# BIBLIOGRAPHIE

1. **ADBS-fr** (*Association des Professionnels de l'Information et de la Documentation*). [On-line]. Rennes (Fr): Comité Réseau des Universités, 1994. Available from internet : <URL : [adbs-info@cru.fr](mailto:adbs-info@cru.fr)>
2. **ADIT**. *Comment chercher une information sur le Web ?* [On-line]. Strasbourg (Fr) : Adit, Juillet 1998 [consulté le 02.07.98]. Available from internet : <URL : <http://www.adit.fr/Recherche/Rech.html>>
3. **AGENTS**. [On-line]. Paris (Fr): Cybion, 1998. Available from internet : <URL : [agent@makelist.com](mailto:agent@makelist.com)>
4. **ANDRIEU, O.** *Méthodes et outils de recherche sur l'Internet*. Paris (Fr): Eyrolles, 1997. 235 p.
5. **BARLOW, L.** *The Spider's Apprentice*. [On-line]. Lexington (USA) : Monash Information Services, 1997 [consulté le 06.07.1998]. Available from internet : <URL : <http://www.monash.com/spidap.html>>
6. **BIBLIO-FR** (*Bibliothécaires et documentalistes francophones*). [On-line]. Rennes: Comité Réseau des Universités, 1993. Available from internet : <URL : [biblio-fr@cru.fr](mailto:biblio-fr@cru.fr)>
7. **BIBLIOSANTE** (*Bibliothécaires et documentalistes en santé*). [On-line]. Mont Saint Aignan (Fr): Crihan, 1998. Available from internet : <URL : [bibliosante@crihan.fr](mailto:bibliosante@crihan.fr)>
8. **CHARRON, J.** *Moteurs de recherche*. [On-line]. Saclay (Fr) : CEA/DIST, Janvier 1997 [consulté le 24.06.1998]. Available from internet : <URL : <http://www-dist.cea.fr/ext/neuf/moteur/>>
9. **CREPUQ**. *Giri 2 : Guide des indispensables de la recherche sur Internet*. [On-line]. Laval (Fr) : Université de Laval, 1997 [consulté le 22.07.98]. Available from internet : <URL : <http://www.bibl.ulaval.ca/vitrine/giri/giri2/tableau.html>>
10. **FELDMAN, S.** Web search services in 1998 : trends and challenges. *Searcher*. [On-line]. 1998, vol 6, n° 6, [consulté le 03.07.1998]. Available from internet : <URL : <http://www.infoday.com/searcher/jun/story2.htm>>
11. **HABIB, D.P. et BALLIOT, R.L.** *How to search the world wide web : a tutorial for beginners and non experts*. [On-line]. Juin 1997 [consulté le 15.08.1998]. Available from internet : <URL : <http://www.ultranet.com/~egrilib/tutor.htm>>

12. **HASSANI.** *Evaluation des méthodes de tri et de classement des documents réponses fournis par les moteurs de recherche.* [On-line]. Lyon (Fr) : Urfist, Juin 1998 [consulté le 13.08.1998]. Available from internet :  
<URL :<http://www.ultranet.com/~egrilib/tutor.htm>>
13. **HOCK, R.** How to do field searching in web search engines : a field tip. *Online.* [On-line]. 1998, vol 22, n° 3 [consulté le 02.07.1998]. Available from internet :  
<URL :<http://www.onlineinc.com/onlinemag/OL1998/hock5.html>>
14. **JOLY, M., BION, N., MERMET, J-M. et al.** *Comment trouver l'information technologique, Guide de l'Ingénieur INSA, chercheur en laboratoire ou en entreprise* [On-line]. Villeurbanne (Fr) : Inst. Nat. Sci. Appl., Doc'INSA, Novembre 1995 [consulté le 08.07.1998]. Available from internet : <URL :[http://www.insa-lyon.fr/Insa/Departements/DocInsa/ISTDOC/istd\\_.html](http://www.insa-lyon.fr/Insa/Departements/DocInsa/ISTDOC/istd_.html)>
15. **LAWRENCE, S. et GILES, C.L.** Searching the World Wide Web. *Science*, 1998, vol 280, p. 98-100.
16. **LIU, J.** *Guide to meta-search engines.* [On-line]. Indiana (USA) : 1998 [consulté le 02.06.1998]. Available from internet :  
<URL :<http://www.indiana.edu/~librcsd/search/meta.html>>
17. **LOMME, L.** L'information électronique en biologie moléculaire. *Documentaliste - Science de l'Information*, 1998, vol 35, n° 3, p. 179-185.
18. **METANEWS.** [On-line]. Paris (Fr): LaMine, 1998. Available from internet : <URL :[metanews@lamine.fr](mailto:metanews@lamine.fr)>
19. **MOTEURS DE RECHERCHE.** [On-line]. Saclay (Fr): CEA, 1998. Available from internet : <URL :[motech@makelist.com](mailto:motech@makelist.com)>
20. **NOTESS, G.R.** On the net : single searching versus megasearching. *Online.* [On-line]. 1998, vol 22, n° 2 [consulté le 02.07.1998]. Available from internet :  
<URL :<http://www.onlineinc.com/onlinemag/OL1998/net3.html>>
21. **NOTESS, G.R.** *Search engine showdown : comparing internet finding tools.* [On-line]. Waterton Glacier (USA) : 1997 [consulté le 02.06.98]. Available from internet : <URL :<http://imt.net/~notess/search/about.htm>>
22. **PLOURDE, J-N.** Définition et application de critères d'évaluation d'outils de recherche dans Internet. *Cursus.* [On-line]. 1995, vol 1, n° 2 [consulté le 02.06.1998].

Available from internet :

<URL :<http://mistral.ere.umontreal.ca/~beaudryg/cursus/vol1no2.plourde.html>>

23. **RANDALL, N. et JONES, D.** *Microsoft Front Page 98*. Le MacMillan. Paris : Simon et Schuster MacMillan, 1997. 925 p.
24. **REVELLI, C.** *Intelligence stratégique sur Internet*. Paris : Dunod, 1998. p. 42-85.
25. **ROSENFELD, L., JANES, J. et VANDER HOLK, M.** *The Internet Compendium : subject guides to health and science resources*. Londres : Mansell, 1995. p. 287-467.
26. **SULLIVAN, D.** *Search Engine Watch* [On-line]. Wesport (USA), 1998 [consulté le 02.07.1998]. Available from internet : <URL :<http://www.searchenginewatch.com>>
27. **THIRION, B. et DARMON, S.** L'internet, les bibliothèques et la médecine. L'expérience du CHU de Rouen. *Documentaliste - Science de l'Information*, 1998, vol 33, n° 3, p. 171-175.
28. **URFIST Alsace Lorraine Franche-Comté.** *Chimie et biologie sur Internet* [On-line]. Strasbourg (Fr) : Urfist, Février 1997 [mise à jour octobre 1997]. Available from internet : <URL :[http://www-scd-ulp.u-strasbg.fr/urfist/internet\\_chimie\\_biologie/principa.htm](http://www-scd-ulp.u-strasbg.fr/urfist/internet_chimie_biologie/principa.htm)>
29. **VEILLE.** [On-line]. Paris (Fr): Cybion, 1998. Available from internet : <URL : [veille@makelist.com](mailto:veille@makelist.com)>
30. **WAYNE, J.** *1997 Healthcare guide to the Internet*. Santa Barbara (USA) : COR Healthcare Resources, 1997. 253 p.

# ANNEXES

# Table des matières

<b>I. LE QUESTIONNAIRE.....</b>	<b>3</b>
A. QUESTIONS GÉNÉRALES.....	3
B. BESOINS EN INFORMATIONS.....	4
C. RECHERCHE D'INFORMATIONS.....	6
<b>II. PRÉSENTATION DES RÉSULTATS DE L'ENQUÊTE.....</b>	<b>8</b>
A. CONNEXION DES UTILISATEURS À INTERNET (QUESTION 1).....	8
B. LANGUES UTILISÉES LORS D'UNE CONSULTATION SUR INTERNET (QUESTION 2).....	8
C. DOMAINES DE RECHERCHE D'INFORMATIONS (QUESTION 3).....	9
D. TYPE D'INFORMATIONS RECHERCHÉES (QUESTION 4).....	9
E. BASES DE DONNÉES CONSULTÉES SUR INTERNET ( SUITE QUESTION 4).....	9
F. CONSULTATION DE LA PAGE D'ACCUEIL ACTUELLE (QUESTION 7).....	13
G. SÉLECTION DES SITES PAR LES UTILISATEURS EN FONCTION DE LEUR DOMAINES D'ACTIVITÉS (QUESTION 8).....	13
H. UTILISATION DES MOTEURS DE RECHERCHE (QUESTION 9).....	14
I. SATISFACTION DES UTILISATEURS PAR RAPPORT À L'UTILISATION DES MOTEURS DE RECHERCHE (QUESTION 10).....	14
J. PRÉFÉRENCE AU NIVEAU DES MOTEURS DE RECHERCHE (QUESTION 11).....	15
K. MODE D'INTERROGATION DES MOTEURS DE RECHERCHE (QUESTION 12).....	16
L. MODE D'INTERROGATION DES MOTEURS DE RECHERCHE (QUESTION 12 BIS).....	16
M. SOURCES D'INFORMATIONS RECHERCHÉES SUR INTERNET (QUESTION 13).....	17
N. MÉTHODOLOGIE DE RECHERCHE SUR INTERNET (QUESTION 14).....	18
O. SUPPORT DE LA MÉTHODOLOGIE (QUESTION 14 BIS).....	18
P. PROFILS PERSONNALISÉS SUR INTERNET (QUESTION 15).....	19
<b>III. LES MOTEURS DE RECHERCHE.....</b>	<b>20</b>
A. PRÉSENTATION DES MOTEURS DE RECHERCHE.....	20
B. PRÉSENTATION DES MOTEURS DE RECHERCHE (SUITE).....	20
C. PRÉSENTATION META MOTEURS DE RECHERCHE.....	20
D. LANGAGE D'INTERROGATION DES MOTEURS DE RECHERCHE.....	20
E. LANGAGE D'INTERROGATION DES META MOTEURS DE RECHERCHE.....	20

## I. Le questionnaire.

Juin 1998

### QUESTIONNAIRE CONCERNANT LES UTILISATEURS D'INTERNET DU SITE DE SANOFI LABEGE

Objectifs : Optimiser l'accès aux ressources d'Internet pour les utilisateurs :

- construction d'un site d'accueil spécifique au centre de recherche de Labège.
- évaluation des outils de recherche présents sur internet et élaboration d'une méthodologie de recherche.

#### A. Questions générales.

	OUI	NON
1. A quelle fréquence vous connectez vous ?  * de temps en temps ? * 1 fois/mois ? * 1 fois/semaine ? * 1 fois/jour ?		
2. Quelle langue utilisez vous principalement pour :  * la consultation : - français - anglais - les deux  * la recherche d'informations: - français - anglais - les deux		

3. Comment obtenez vous les adresses des sites que vous venez consulter (presse, collègue, ...) ?

## B. Besoins en informations.

	OUI	NON
<p>4. Dans le cadre de vos activités, dans quel(s) domaine(s) cherchez vous de l'information ?</p> <ul style="list-style-type: none"> <li>* chimie</li> <li>* biochimie</li> <li>* biologie moléculaire</li> <li>* microbiologie</li> <li>* biotechnologie</li> <li>* immunologie</li> <li>* pharmacologie</li> <li>* toxicologie</li> <li>* neurobiologie</li> <li>* pathologie humaine</li> <li>* analyse</li> <li>* techniques de laboratoires / protocoles</li> <li>* documentation</li> <li>* informatique</li> <li>* sécurité</li> <li>* développement clinique</li> <li>* affaires réglementaires</li> <li>* assurance qualité</li> <li>* autres</li> </ul>		
<p>5. Par rapport aux domaines précédemment cités, de quels types d'informations avez vous besoin :</p> <ul style="list-style-type: none"> <li>* sites de vulgarisation</li> <li>* généralités sur un domaine (ex : cours, ...)</li> <li>* articles de journaux presse généraliste</li> <li>* articles de journaux presse spécialisée</li> <li>* brevets (ex : EPO, IBM Patent Server)</li> <li>* sites « ressources » (ex : Pasteur)</li> <li>* colloques, congrès, salons professionnels,...</li> <li>* formation continue</li> <li>* base de données : <ul style="list-style-type: none"> <li>séquences de gènes</li> <li>séquences de protéines</li> <li>données cristallographiques</li> <li>composés chimiques</li> <li>références bibliographiques</li> <li>nomenclature</li> </ul> </li> </ul>		

<p style="text-align: center;">autres</p> <p>* adresses électroniques * sites de fournisseurs</p>		
<p>6. Utilisez vous les adresses de la page d'accueil actuelle ?</p> <p>Remarques :</p>		
<p>7. Quels sont les domaines pour lesquels vous avez du mal à trouver de l'information ?</p>	/	/
<p>8. Quels sont les types d'informations pour lesquels vous avez du mal à trouver de l'information ?</p>	/	/
<p>9. Y a t-il des sites que vous consultez régulièrement ? Lesquels ?</p>		
<p>10. Seriez vous d'accord pour collaborer à la sélection des sites, en rapport avec vos activités, destinée à l'élaboration de la page d'accueil ?</p>		

### C. Recherche d'Informations.

	OUI	NON
<p>11. Avez vous déjà recherché des informations par des moteurs de recherche ?</p> <p>Si oui, lesquels ?</p>		
<p>12. Etes vous satisfait des réponses obtenues par les moteurs précédemment cités ?</p> <p>Remarques :</p>		
<p>13. Préférez vous les moteurs :</p> <p>* « thématiques » (type Yahoo) ? Lesquels ?</p> <p>* « généralistes » (type Alta Vista) ? Lesquels ?</p> <p>* méta-moteurs ? Lesquels ?</p>		
<p>14. Lorsque vous interrogez, est ce par :</p> <p>* mots (sans autres précisions)?</p> <p>* mots dans les URLs ?</p> <p>* mots dans les titres ?</p> <p>* langage naturel ?</p>		
<p>15. Lorsque vous interrogez, est ce par :</p> <p>* mots-clés ?</p> <p>* concept ?</p>		

* les deux ?		
16. Quelles sources d'infos vous intéressent ? * sites Web ? * les listes de discussion ? * les deux ?		
17. Seriez vous intéressé par une méthodologie de recherche sur Internet ?  Si oui : * en ligne ? * en guide papier ? * les deux ?		
18. A terme, seriez vous intéressé par la mise en place de profils personnalisés sur Internet?		

19. En résumé, quel est selon vous « une séance idéale sur Internet » :

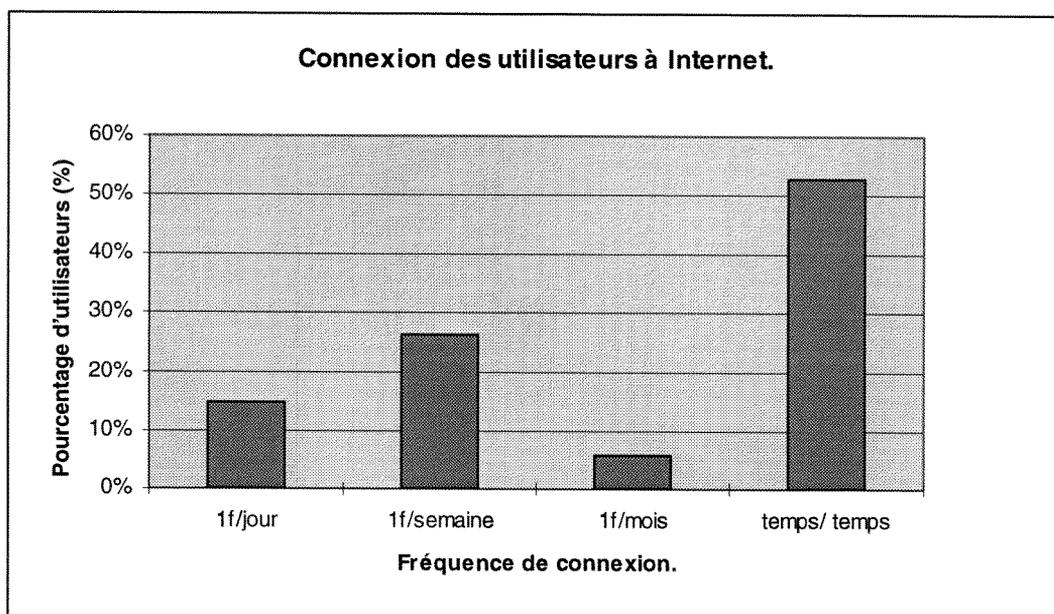
\* au niveau du matériel ?

\* au niveau de l'accompagnement humain ?

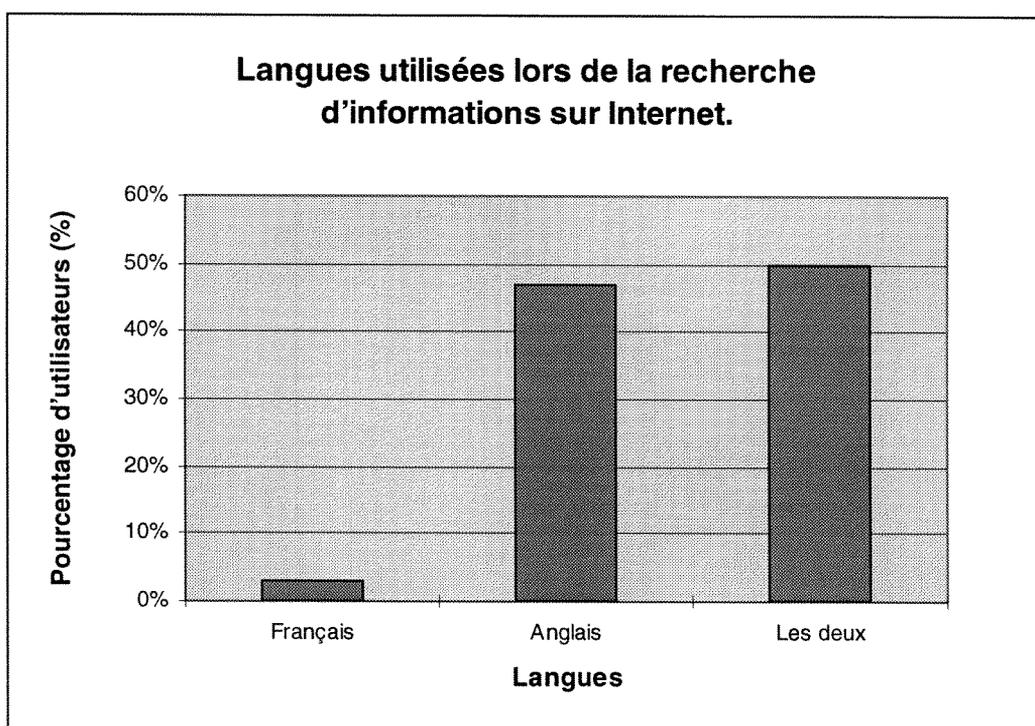
\* au niveau du site d'accueil ?

## II. Présentation des résultats de l'enquête.

### A. Connexion des utilisateurs à Internet (question 1)



### B. Langues utilisées lors d'une consultation sur Internet (question 2)



**C. Domaines de recherche d'informations (question 3)**

voir le graphique page 10.

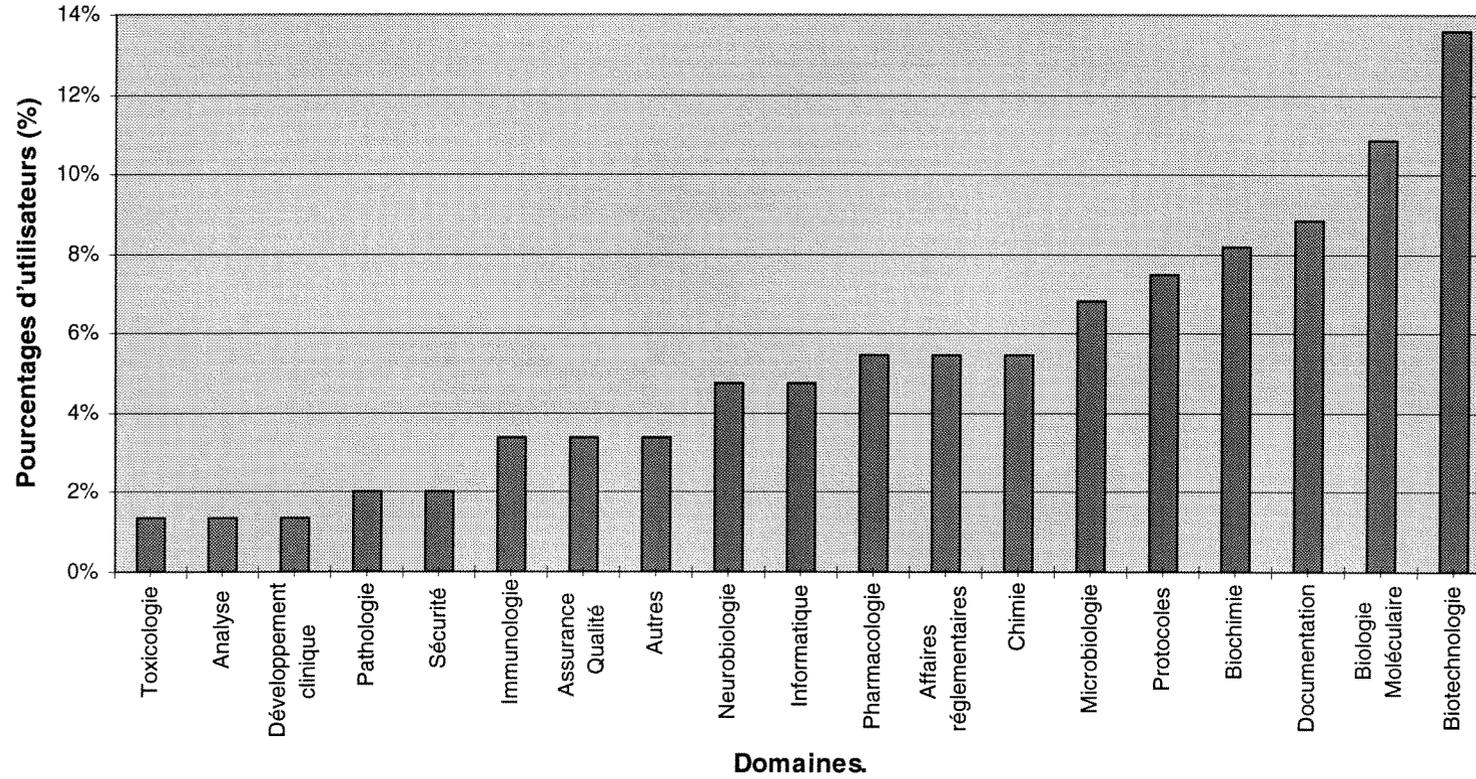
**D. Type d'informations recherchées (question 4)**

voir le graphique page 11.

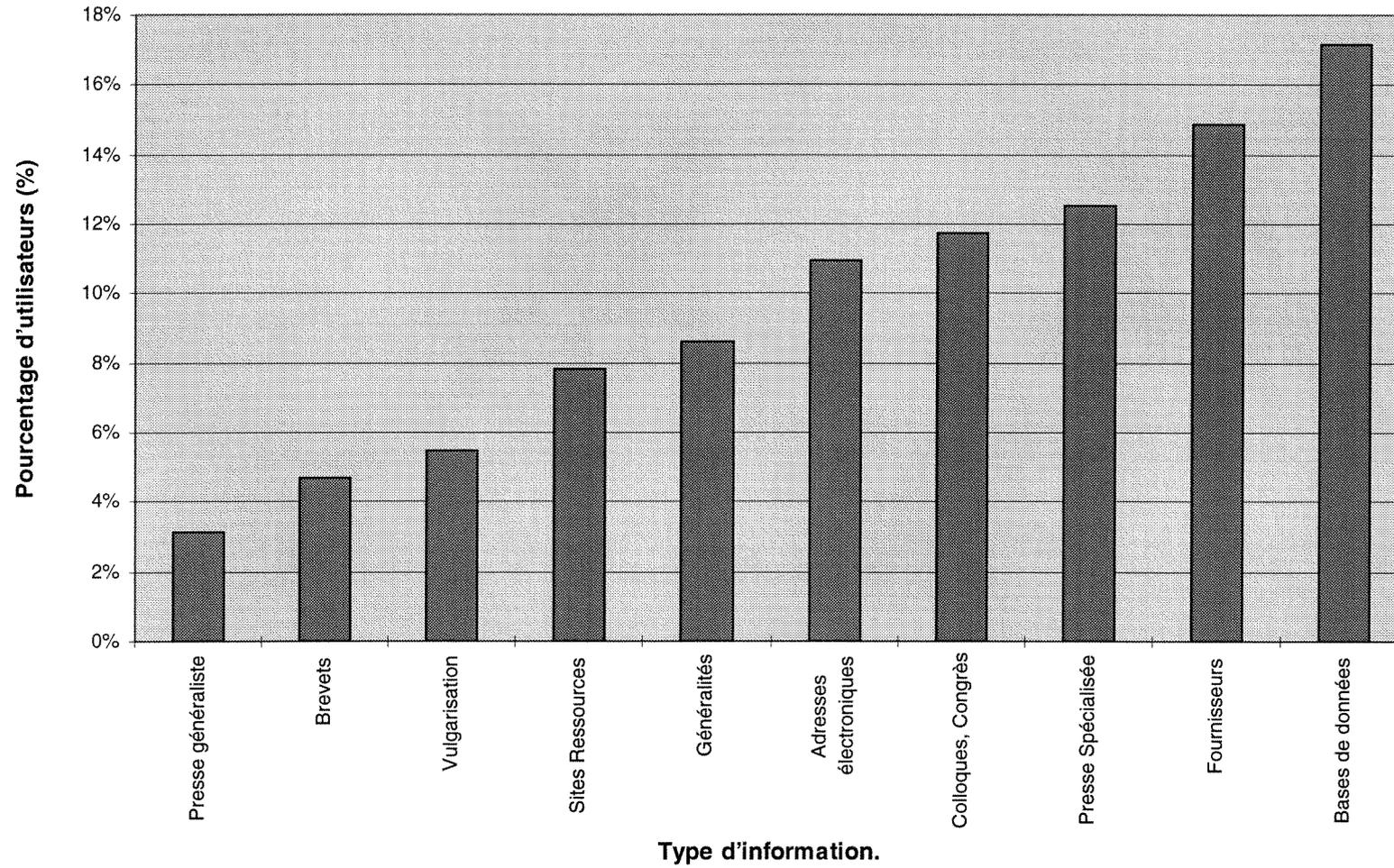
**E. Bases de données consultées sur Internet ( suite question 4)**

voir le graphique page 12.

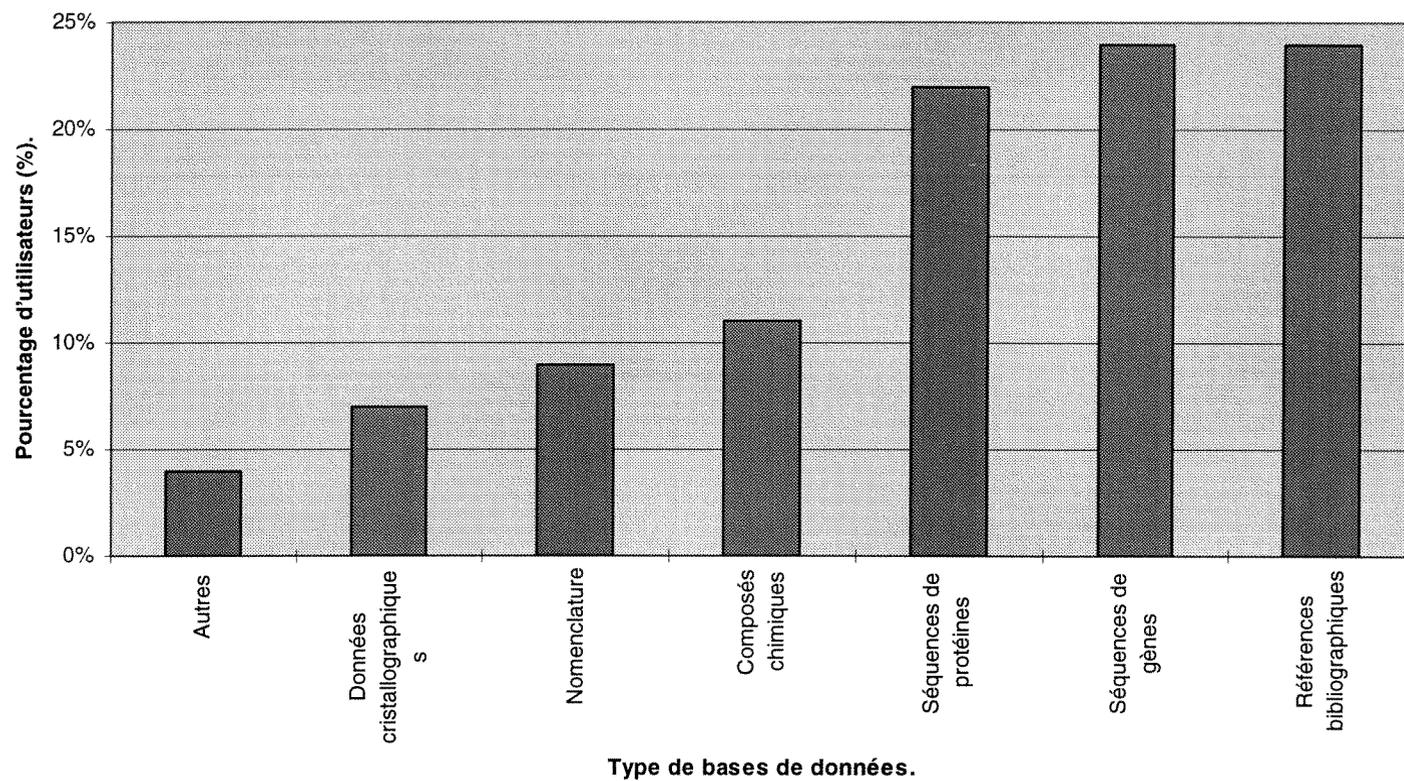
### Domaines de recherche d'informations.



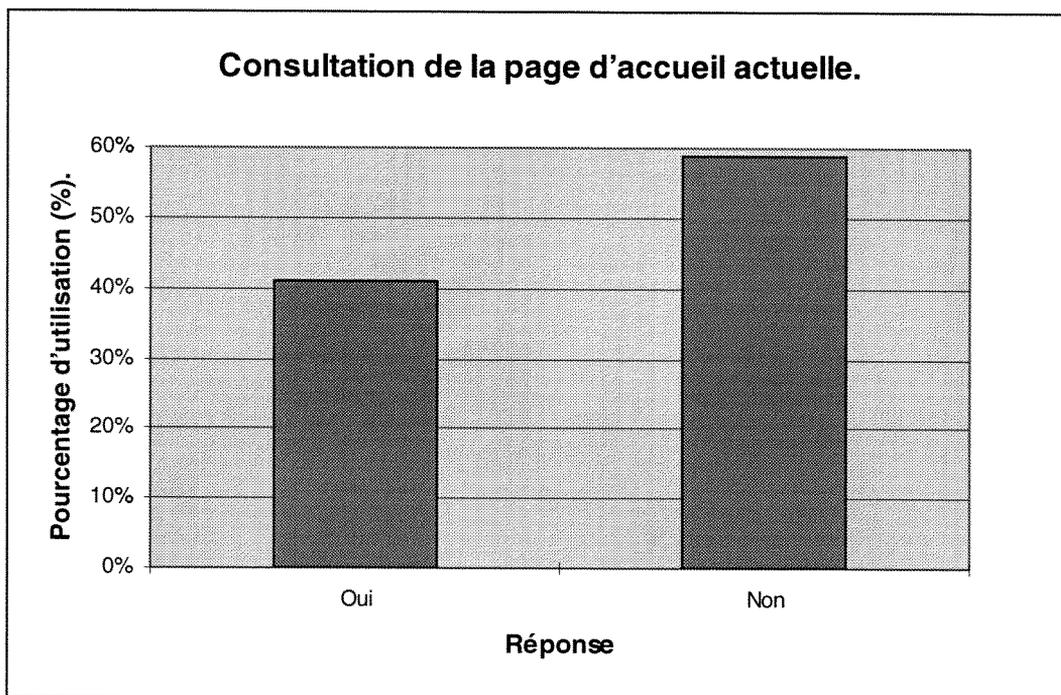
### Type d'informations recherché.



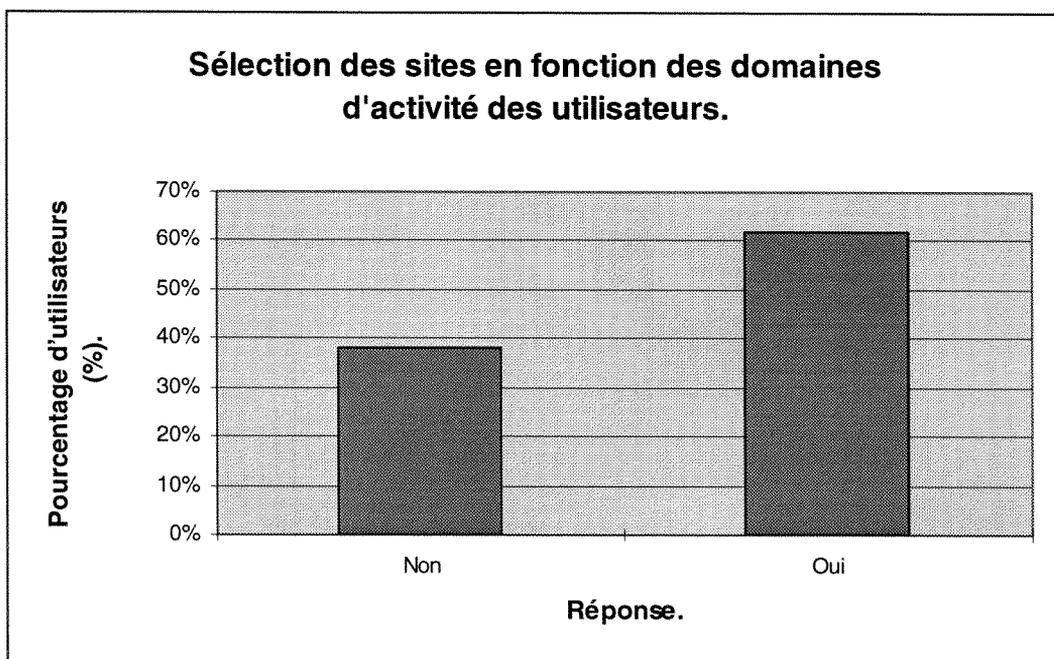
### Bases de données consultées sur Internet.



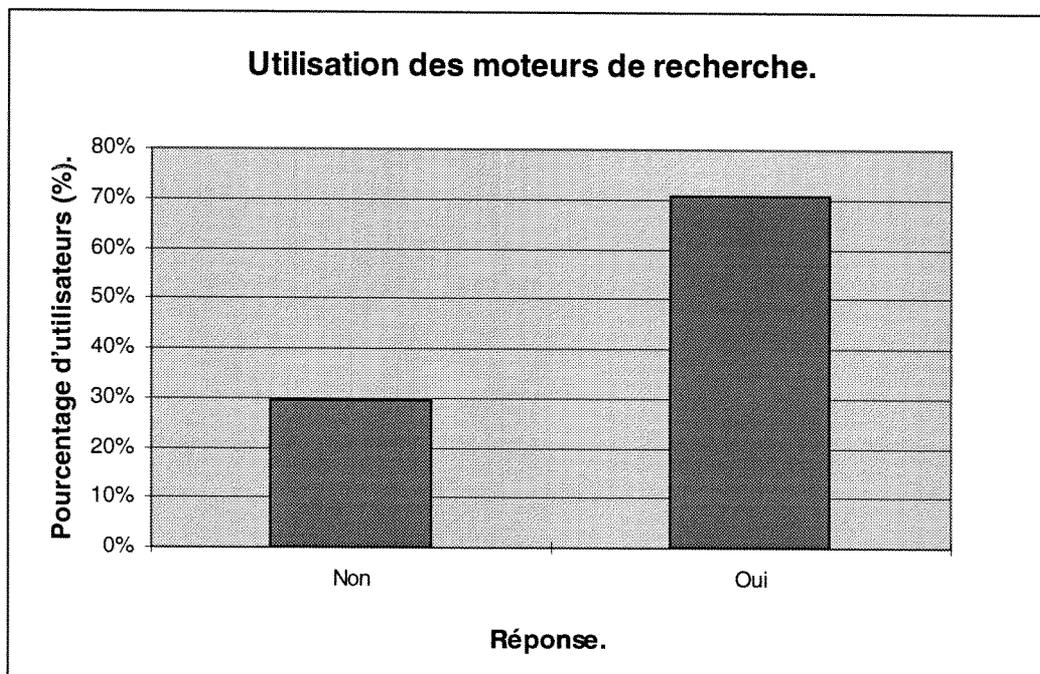
**F. Consultation de la page d'accueil actuelle (question 7)**



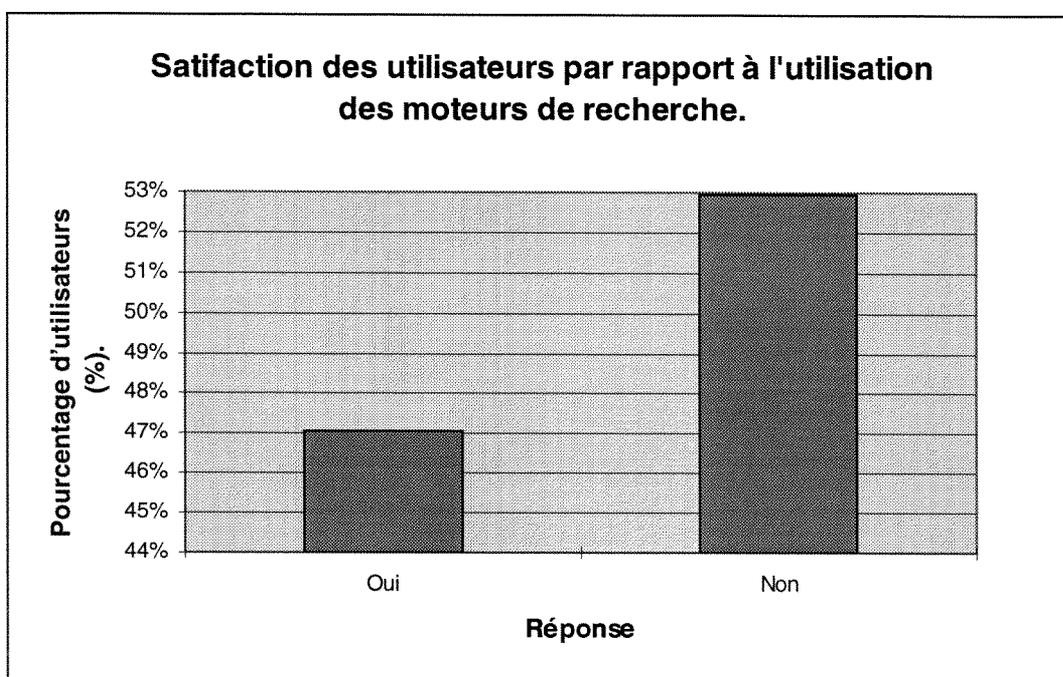
**G. Sélection des sites par les utilisateurs en fonction de leur domaines d'activités (question 8)**



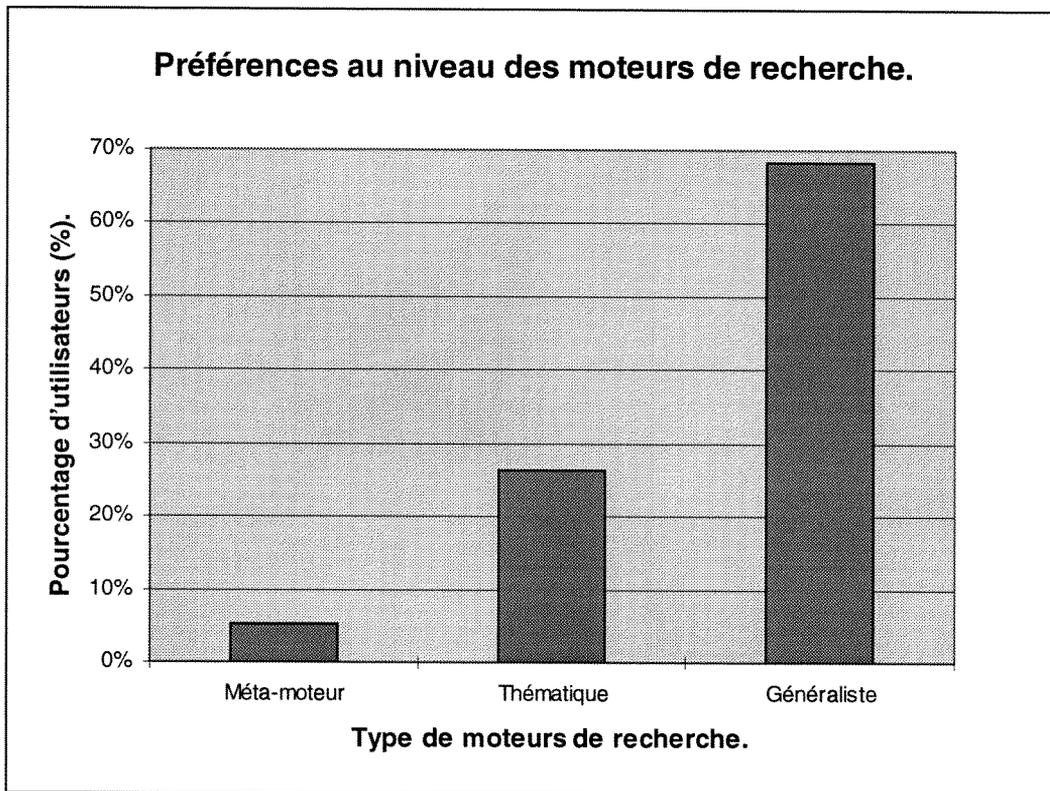
## H. Utilisation des moteurs de recherche (question 9)



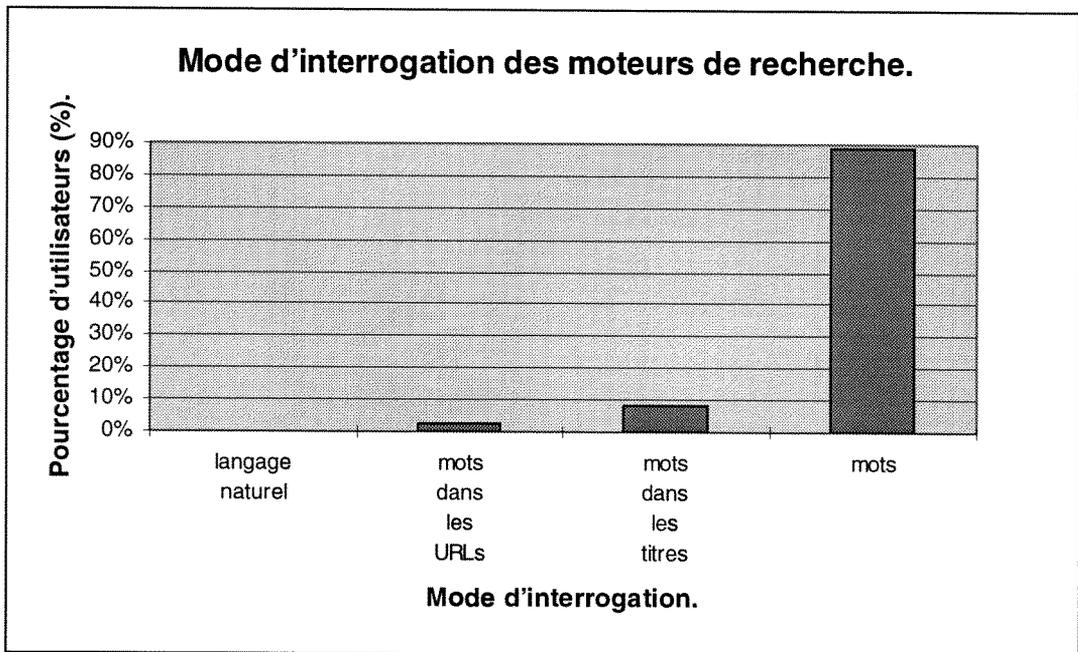
## I. Satisfaction des utilisateurs par rapport à l'utilisation des moteurs de recherche (question 10)



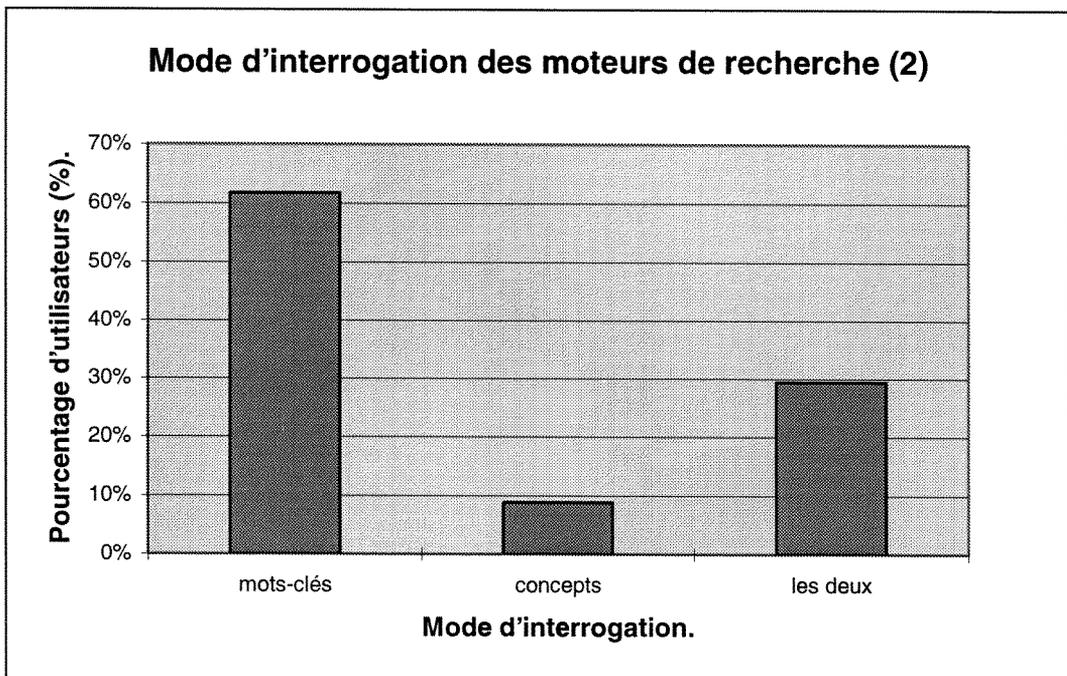
**J. Préférence au niveau des moteurs de recherche (question 11)**



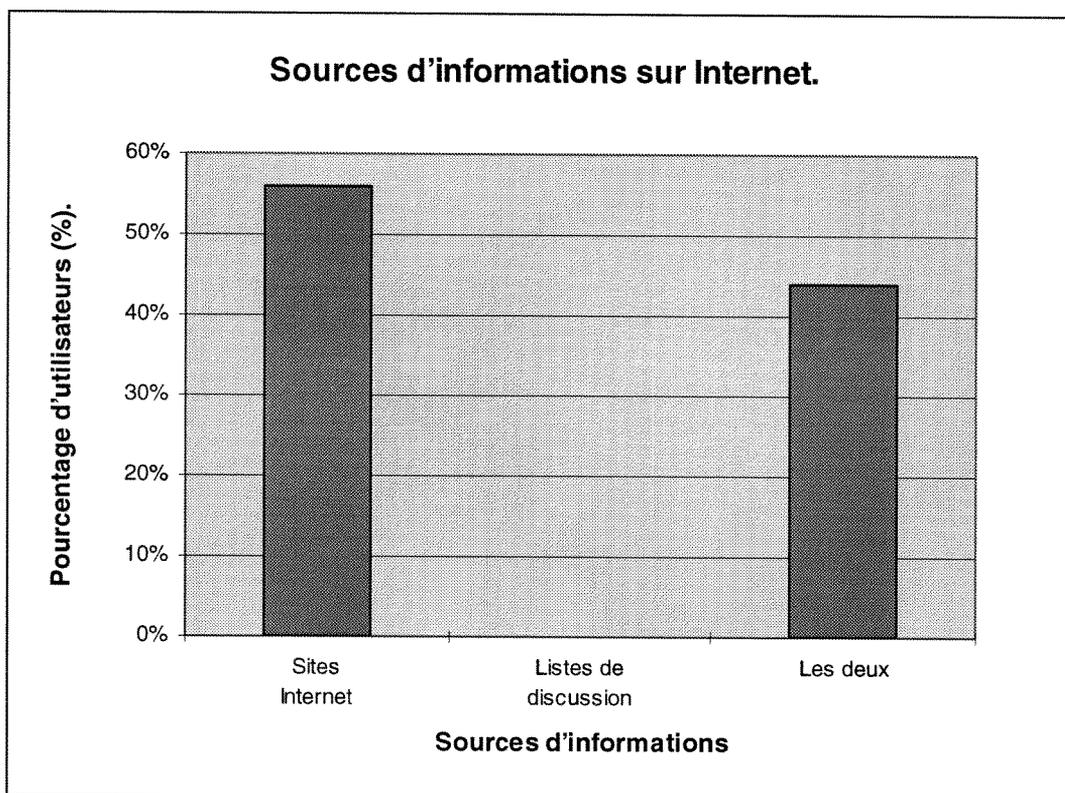
**K. Mode d'interrogation des moteurs de recherche (question 12)**



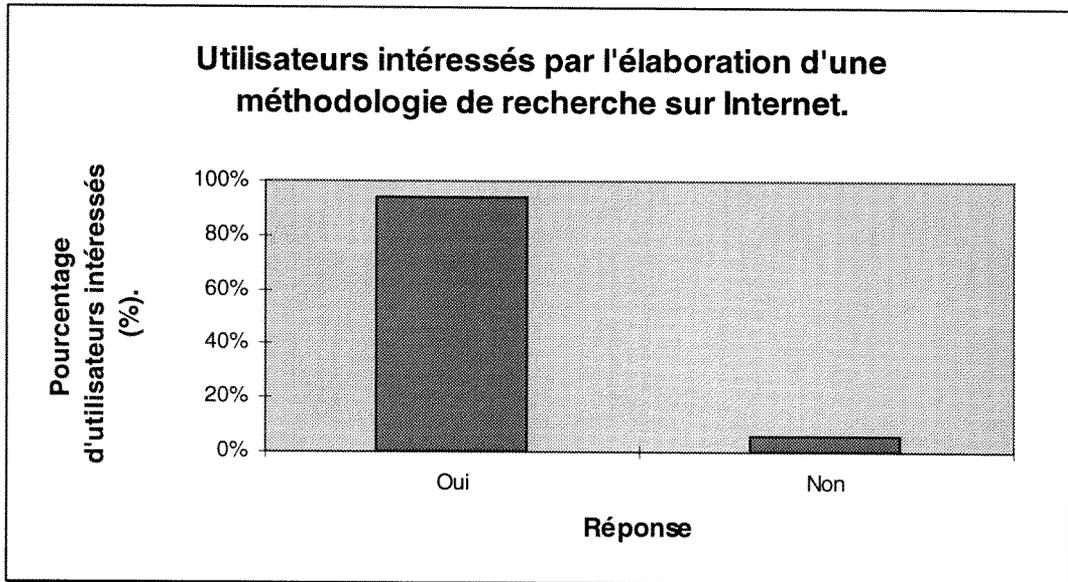
**L. Mode d'interrogation des moteurs de recherche (question 12 bis)**



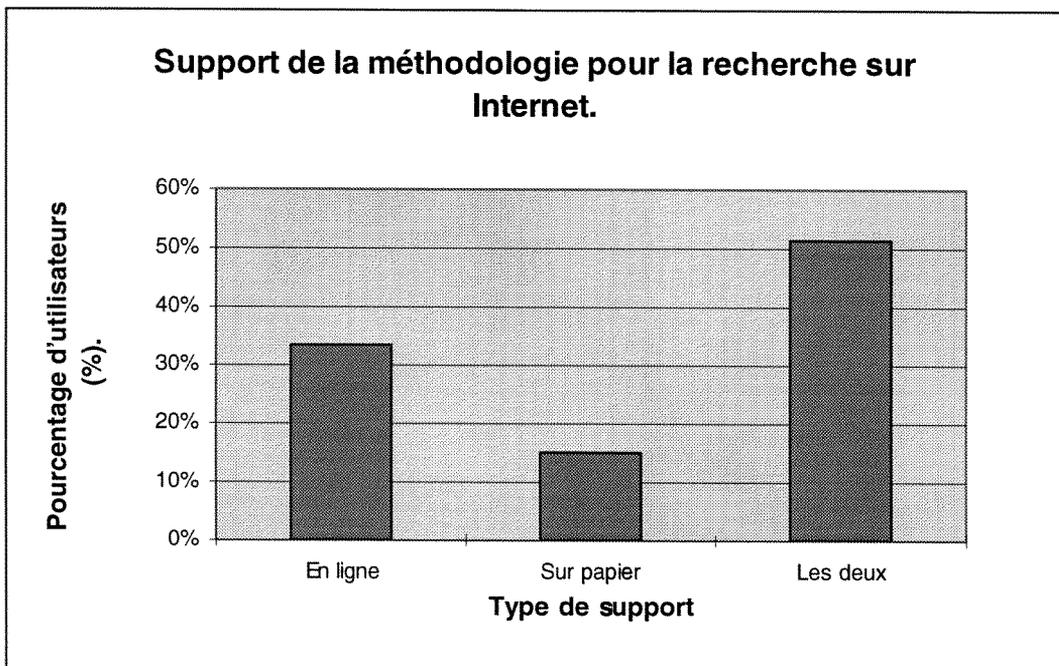
**M. Sources d'informations recherchées sur Internet (question 13)**



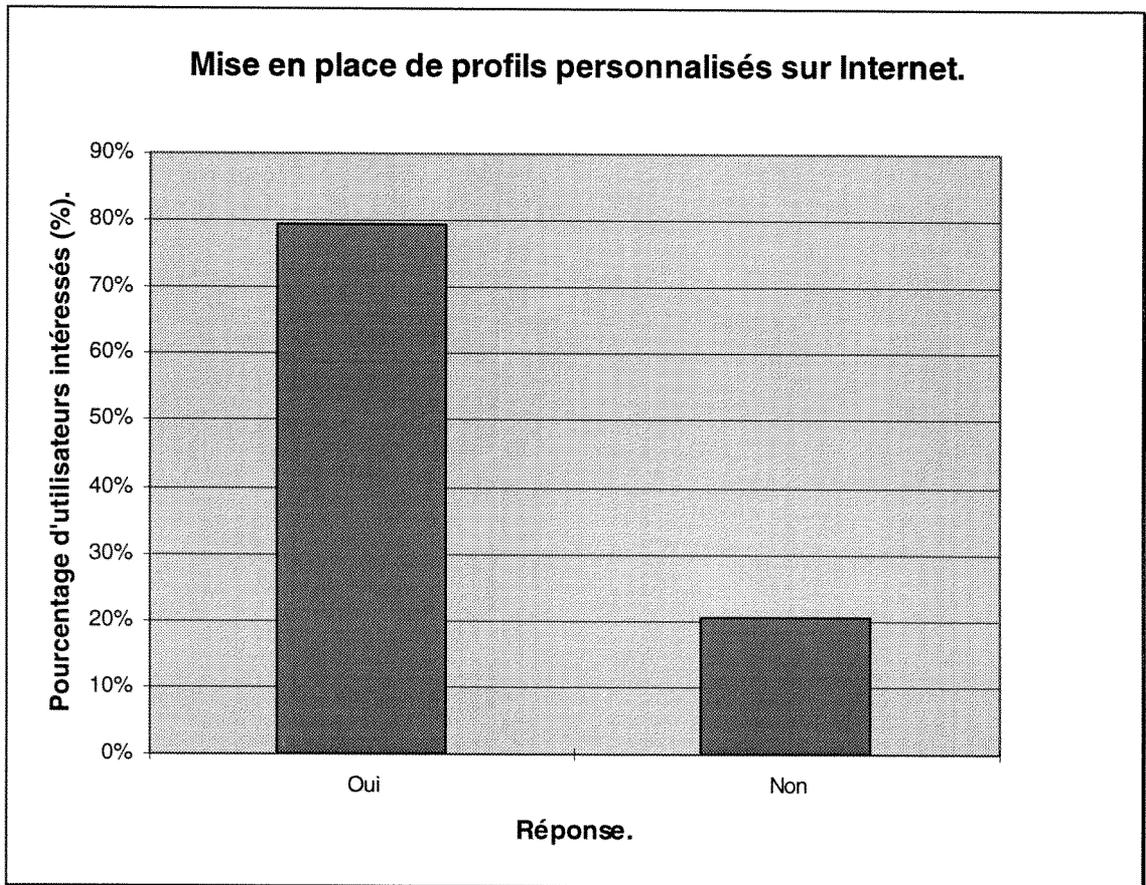
**N. Méthodologie de recherche sur Internet (question 14)**



**O. Support de la méthodologie (question 14 bis)**



**P. Profils personnalisés sur Internet (question 15)**



### **III. LES MOTEURS DE RECHERCHE.**

Ces présentations ont été établies à partir de guides (Charron, 1997 ; Sullivan, 1998) mais surtout à partir des sites des moteurs de recherche.

#### **A. Présentation des moteurs de recherche.**

Tableau page 21.

#### **B. Présentation des moteurs de recherche (suite).**

Tableau page 22.

#### **C. Présentation meta moteurs de recherche.**

Tableau page 23.

#### **D. Langage d'interrogation des moteurs de recherche.**

Tableau page 24.

#### **E. Langage d'interrogation des meta moteurs de recherche.**

Tableau page 25.

## PRESENTATION DES MOTEURS DE RECHERCHE.

	<b>AltaVista</b>	<b>Excite</b>	<b>Galaxy</b>	<b>HotBot</b>
<b>Adresse</b>	www.altavista.digital.com	www.excite.com	galaxy.einet.net	www.hotbot.com
<b>Type de moteur</b>	index	index et répertoire	index et répertoire	index
<b>Volume de la base (pages)</b>	140 000 000	55 000 000	350 000 liens	110 000 000
<b>Fréquence de mise à jour</b>	1 jour à 1 mois	1 à 3 semaines	?	1 jour à 2 semaines
<b>Type de fichiers collectés</b>	WWW, Usenet , News	WWW, telnet	WWW, gopher, telnet	WWW, usenet, multimédia
<b>Couverture</b>	mondiale	mondiale	?	mondial
<b>Sujet</b>	général	général	général	général
<b>Fréquence de visite des documents</b>	?	hebdomadaire	?	hebdomadaire
<b>Données indexées</b>	titres, URLs, texte URLs	titres, sous-titres, texte URLs, résumé, texte intégral	titres, sous-titres, taille et date de fichiers, URLs, texte URLs, extraits	méta-tags, date et taille de fichiers, texte intégral
<b>Méthodes d'élimination des mots vides</b>	Fréquence d'apparition des mots (si le terme apparaît trop souvent, il sera éliminé car considéré comme vide de sens)	Fréquence d'apparition des mots (si le terme apparaît trop souvent, il sera éliminé car considéré comme vide de sens)	Les mots les plus communs de l'anglais sont retirés.	/
<b>Type de question</b>	booléen, liste de mots	booléen, liste de mots	booléen	booléen, liste de mots
<b>Champs recherchés</b>	texte intégral	texte intégral	texte intégral	texte intégral
<b>Traitement manuel</b>	/	catalogage	/	/
<b>Particularité</b>		recherche par concept	un des meilleurs outils de type répertoire	

## PRESENTATION DES MOTEURS DE RECHERCHE (suite)

	<b>Infoseek</b>	<b>Lycos</b>	<b>Magellan</b>	<b>Yahoo</b>
<b>Adresse</b>	www.infoseek.com	www.lycos.fr	www.mckinley.com	www.yahoo.fr
<b>Type de moteur</b>		index	index et répertoire	index et répertoire
<b>Volume de la base</b>	30 000 000 WWW	30 000 000	4 000 000 ( ? )	40 000
<b>Fréquence de mise à jour</b>	1 jour à 2 mois	1 à 2 semaines	mensuel ( ? )	journalière ( ? )
<b>Type de fichiers collectés</b>	WWW, gopher, FTP, usenet, CSA Biomedical database, ...	WWW, gopher, FTP, multimédia, graphiques, programmes	WWW, gopher, telnet, ftp, telnet	WWW, Reuter
<b>Couverture</b>	mondiale	mondiale	mondiale	française
<b>Sujet</b>	général	général	général	général
<b>Fréquence de visite des documents</b>	20 fois/mois	mensuel	?	?
<b>Données indexées</b>	titres, sous-titres, meta-tags, taille fichier, extraits et texte intégral	titres, sous-titres, date et taille du fichier, URLs, texte URLs, résumé, extrait	titres, URLs, texte URLs, résumés	?
<b>Méthodes d'élimination des mots vides</b>	Les mots les plus communs de l'anglais sont retirés.	Fréquence d'apparition des mots (si le terme apparaît trop souvent, il sera éliminé car considéré comme vide de sens)	Les mots les plus communs de l'anglais sont retirés.	/
<b>Type de question</b>	langage naturel, liste de mots	booléen, liste de mots	liste mots	?
<b>Champs recherchés</b>	URL, titre, texte intégral, URL citée		texte intégral	?
<b>Traitement manuel</b>	/		indexation et résumés	indexation
<b>Particularité</b>		Pour le repérage d'informations		

## PRESENTATION DES META MOTEURS DE RECHERCHE

	<b>Profusion</b>	<b>Metacrawler</b>	<b>Dogpile</b>
<b>Type de moteur</b>	méta-moteur	méta-moteur	méta-moteur
<b>Nombre de moteurs utilisés</b>	9	9	13 WWW 4 usenet 2 FTP business news
<b>Elimination des doublons</b>	O	O	O
<b>Traduction des requêtes aux différents moteurs</b>	O	?	O
<b>Utilisation d'opérateurs booléens</b>	O	O	O

## LANGAGE D'INTERROGATION DES MOTEURS DE RECHERCHE.

	<b>AltaVista</b>	<b>Excite</b>	<b>HotBot</b>	<b>Infoseek</b>	<b>Lycos</b>	<b>Magellan</b>
<b>Opérateurs booléens</b>	and, or, and not, near système +/-	and, or, and not système +/-	and, or, not système +/-	and, or, not, near système +/-	and, or, not système +/-	and, or, not
<b>Opérateurs de proximité</b>	O	N	N	O	N	N
<b>Opérateurs d'adjacence</b>	O	O	O	O	N	N
<b>Parenthésage</b>	O	O	N	N	N	N
<b>Troncature automatique</b>	N	lemmatisation et troncature à droite	N	N	lemmatisation, troncature désactivable	N
<b>Troncature manuelle</b>	droite (3 lettres), interne	N	N	N	N	N
<b>Champs de recherche</b>	texte intégral	texte intégral	texte intégral	URL, titre, texte intégral, URL cité	URL, titre, résumé, URL cité	texte intégral
<b>Champs spécifiables</b>	URL, titres, url cités	/	URL	/	/	/
<b>Gestion des Maj/min</b>	O	O	N	O	O	minuscule seule
<b>Gestion de l'accentuation</b>	O	N	N	N	N	N
<b>Phrase (“ ”)</b>	Oui	Oui	Non	Oui	Oui	Non

## LANGAGE D'INTERROGATION DES META MOTEURS DE RECHERCHE.

	<b>Profusion</b>	<b>Metacrawler</b>	<b>Dogpile</b>
<b>Opérateurs booléens</b>	AND, OR, NOT, NEAR	and, or, phrase système +/- (pas de mélange d'opérateurs)	AND, OR, NOT, NEAR
<b>Parenthésage</b>	O	N	Oui
<b>Troncature automatique</b>	?	N	?
<b>Phrase</b>	Oui si le moteur l'utilise	“ ”	Oui si le moteur l'utilise

