

Xavier Ciana

La Recherche d'information sur Internet :

théorie et diffusion des résultats, un exemple d'application en PHP

**Travail présenté à la Ecole d'Information Documentaire
pour l'obtention du diplôme**

Genève

2000

CIANA, Xavier

La recherche d'information sur Internet : théorie et diffusion des résultats, un exemple d'application en PHP / Xavier Ciana. – Genève : E. I. D., 2000. – II, 63 p. : ill. ; 30 cm

Résumé

Le présent travail est structuré autour de deux pôles:

- Un premier pôle théorique, traitant dans sa plus grande partie de la recherche d'information sur le web - ses possibilités et ses limites. Cette partie théorique traite également des techniques utilisées à l'heure actuelle pour le développement de sites web.
- Un second pôle, plus pratique, présentant les démarches effectuées et la méthodologie utilisée afin de réaliser le mandat suivant : créer un répertoire de sites Internet concernant l'emploi en Suisse romande.

Ce travail a été dirigé par Mme Anne-Christine Gourdal, assistante à l'Ecole d'Information Documentaire, supervisé par Mme Ariane Rezzonico, responsable de formation à l'Ecole d'Information Documentaire et commandité par le cabinet Daniel Porot.

Les propos émis dans ce travail n'engagent que la responsabilité de l'auteur.

Remerciements

Je tiens à remercier très chaleureusement les personnes suivantes qui m'ont apporté leur aide et leurs conseils dans la réalisation de ce travail de diplôme :

Anne-Christine Gourdal

Marie-Thé Ciana
Pierre Crevoisier
Nicolas Derron
Julien Notter
Daniel Porot
Ariane Rezzonico
François Rudhart

Résumé	2
Remerciements	3
1 Introduction	5
2 Concepts théoriques	6
2.1 La Recherche d'information sur le web	7
2.1.1 Spécificité de l'information sur le web	8
2.1.2 Le père Noël n'existe pas	9
2.1.3 Les outils de recherche	10
2.1.4 Annuaire	10
2.1.5 Moteurs de recherche	13
2.1.6 Meta-moteurs on-line	17
2.1.7 Agents intelligents	19
2.1.8 Outils spécifiques	20
2.1.9 Trouver ce que l'on cherche ou la recherche d'une information pertinente	21
2.1.10 Tendances et perspectives d'avenir	23
2.1.11 Les stratégies de recherche	25
2.1.12 Les limites de la recherche : Le web invisible	27
2.2 Les portails	28
2.2.1 La preuve vivante : Yahoo!	28
2.2.2 Les différents type de portail	28
2.2.3 Les différentes fonctions d'un portail	30
2.3 Création de site web	32
2.3.1 Le web : communication et interaction	32
2.3.2 Les technologies du web	32
2.3.3 Pages statiques et pages dynamiques	33
2.3.4 Pourquoi utiliser des pages dynamiques ?	36
2.3.5 Les langages de programmation de pages dynamiques : un aperçu	37
2.4 Le web et l'emploi	38
2.4.1 Les nouveaux outils de l'emploi	38
2.4.2 Le web et l'emploi en Suisse romande	40
3 Réalisation du projet	41
3.1 Genèse et objectif du projet	42
3.2 Le mandat	43
3.3 Méthodologie de travail	45
3.3.1 Méthodes de recherche utilisées	45
3.3.2 Résultats des recherches	48
3.3.3 Évaluation et sélection des sites	49
3.3.4 Structuration des résultats : L'arborescence du répertoire	50
3.4 Réalisation technique	51
3.4.1 La solution technique choisie : phpHoo 2	53
3.4.2 La mise en place de l'application	55
3.5 Utilisation et évolution future du répertoire	56
4 Conclusion	57
5 Bibliographie	59
6 Lexique	64

1 Introduction

A l'origine de ce travail de diplôme se trouve un projet du cabinet Daniel Porot : la " Boutique Emploi ". L'idée de ce projet étant la création d'un lieu dédié à l'emploi, - la Boutique Emploi -, qui servirait de plate-forme de rencontre entre les entreprises et les personnes à la recherche d'un emploi. Ce lieu offrirait également une palette de services et d'informations aux personnes à la recherche d'un travail.

C'est dans le contexte de réalisation de ce projet que j'ai rencontré Monsieur Daniel Porot. Le mandat qu'il m'a proposé est le suivant : réaliser un site Internet pour la Boutique Emploi : un lieu virtuel réunissant les ressources disponibles sur Internet (et plus particulièrement en suisse romande) susceptibles d'intéresser les personnes à la recherche d'un emploi, ainsi qu'aux entreprises. En substance, réunir un contenu et créer un outil (un site web en l'occurrence) qui soit utilisé à la fois par des candidats et des entreprises.

L'idée étant de créer, à partir de la source brute d'information qu'est Internet, en un outil utilisable, ceci en réalisant un tri et une présentation claire des ressources offertes par Internet. En d'autres termes, de créer une porte d'entrée de l'emploi sur le web suisse romand. Cette porte d'entrée prenant la forme d'un annuaire ou d'un répertoire de sites.

Je trouvais ce concept intéressant et il constituait pour moi le support idéal à la réalisation d'un projet touchant à des domaines auxquels je porte un intérêt marqué. A savoir : la recherche sur Internet et la réalisation d'un produit documentaire grâce à l'utilisation des nouvelles technologies de l'information.

Mon mémoire est structuré de la manière suivante :

- Dans une première partie de taille importante, j'expose les concepts théoriques auxquels je me suis intéressé et que j'ai été amené à utiliser tout au long de la réalisation de ce travail de diplôme. Tous n'ont pas eu d'utilité directe et concrète dans la réalisation pratique de mon travail, mais ces concepts théoriques représentent l'état de mes connaissances dans les domaines touchant à ce travail de diplôme.
- Dans une seconde partie plus pratique, je présente le projet, son contexte, et sa réalisation. J'expose également ma méthodologie de travail ainsi que les choix et démarches effectués pour remplir mon mandat.

2 Concepts théoriques

2.1 La Recherche d'information sur le web

Selon une étude¹ de Cyveillance – une société spécialisée dans le e-business – datant du 10 juillet 2000, le web aurait maintenant atteint la taille de 2,1 milliards de pages. Son taux de croissance actuel serait de 7 millions de pages supplémentaire chaque jour. Selon une autre étude du groupe IDG (un groupe spécialiste des médias), d'ici 2003 le web devrait avoir atteint 13 milliards de pages. Ces chiffres vertigineux nous aident à saisir la difficulté et la complexité d'une recherche d'information sur le web. La toile est devenue le plus grand réservoir d'informations au monde, et corollaire inévitable de sa croissance exponentielle, une base d'informations extrêmement hétérogène et inorganisée. Ce chapitre traite de cet exercice parfois périlleux qu'est la recherche d'information sur le web : ses outils, ses méthodes et ses limites.

Il est tout d'abord utile de préciser qu'Internet n'est pas synonyme du web. Le réseau Internet est composé de plusieurs " sphères " et le web ne représente qu'une de celles-ci. Parmi ces autres sphères d'Internet (indépendantes du web), on citera :

- **FTP (File Transfert Protocol)**
FTP est le protocole utilisé pour l'échange de fichiers. Il existe des millions de serveurs FTP publics ou privés connectés à Internet et représentant une base titanesque de données et de logiciels.
- **Telnet**
Protocole utilisé afin de se connecter à une machine distante. Telnet est souvent utilisé pour consulter des bases de données (des catalogues de bibliothèques). Cependant les interrogations via Telnet ont tendance à disparaître au profit d'interfaces web.
- **Gopher**
Gopher est une sorte d'ancêtre du web, remplissant les même fonctions, mais c'est aujourd'hui une technologie complètement obsolète. La quasi-totalité des applications Gopher ont été supplantées par le web.
- **IRC (Internet Relay Chat)**
IRC est un réseau de serveurs permettant le dialogue en temps réel.
- **Les newsgroups**
Les newsgroups sont les listes de discussions existantes sur Internet. Ces listes de discussions sont relayées par un réseau de serveurs appelé "usenet".
- **E-Mail**
L'application la plus répandue est sans aucun doute le courrier électronique.

Tous ces différents services Internet requièrent des logiciels spécifiques qui permettent l'accès et la lecture des différents types d'informations, on n'utilisera pas le même programme pour lire les messages d'une liste de discussion, pour transférer des fichiers ou pour " chatter " via IRC.

Toutefois, on assiste maintenant à un décloisonnement de ces différentes sphères, qui étaient jusqu'ici indépendantes l'une de l'autre. La technologie du web (et particulièrement les logiciels utilisés pour surfer) a contribué à établir des passerelles entre les différentes sphères, elle fait office de plate-forme commune. Il est actuellement possible d'échanger des fichiers en utilisant le protocole FTP ou de discuter sur IRC via un

¹ <http://www.cyveillance.com/newsroom/pressr/000710.asp>

navigateur web classique (Internet Explorer, Netscape Navigator). Certains sites offrent également la possibilité d'accéder aux newsgroups² au travers de leurs pages web. Suite à ce décloisonnement, il est pratiquement possible d'effectuer des recherches dans toutes les différentes sphères d'Internet en utilisant un navigateur web. L'expression " recherche d'information sur le web " est donc *pratiquement* synonyme à recherche d'information sur Internet. Ceci d'autant plus que certaines sphères du net représentent une source d'information assez négligeable, notamment IRC (les discussions qui s'y tiennent sont éphémères et rarement informatives) et Gopher (du fait de son obsolescence).

Ceci explique donc le titre du chapitre, qui n'est pas " recherche d'information sur Internet ", mais bien " recherche d'information sur le web ", la partie d'Internet non accessible à partir de la toile n'étant pas traitée ici.

2.1.1 Spécificité de l'information sur le web

Le web a bouleversé les schémas traditionnels de diffusion de l'information, il s'agit du premier média réellement global, et c'est le seul media permettant à chacun de s'exprimer de manière libre (avec il est vrai, plus ou moins de chance d'être entendu). Si dans les faits le web est une réelle révolution médiatique, une évolution dont nous ne sommes pas prêt de voir la forme définitive, les mentalités quant à elles n'évoluent pas aussi rapidement que la technologie. Qu'en est-il de notre manière d'appréhender l'information ? Ne continue-t-on pas à se comporter face à des informations provenant du web, comme face à des informations issues de source traditionnelle, auxquelles nous sommes habituées ? Il est salutaire d'avoir un certain recul, d'effectuer une recontextualisation face aux informations provenant du web. Pour réaliser cela, il est bon de formaliser un certain nombre de caractéristiques de l'information diffusée grâce au web.

L'information disponible sur le web possède les caractéristiques suivantes :

- En pleine expansion :
2,1 milliards de pages pour l'année 2000 et probablement treize fois plus d'ici trois ans.
- Hétérogène:
 - **Au niveau de son contenu** : n'importe qui peut prendre l'initiative de diffuser de l'information, des personnes privées, aussi bien que des entreprises. L'information peut donc revêtir un caractère public et/ou commercial. L'information est également internationale, multilingue et multiscript.
 - **Au niveau de sa forme** : cette information peut prendre toutes les formes : extrait ou intégralité d'articles ou de monographies publiées auparavant sur support papier, textes inédits de toute nature, base de données, documents photographiques, sonores, vidéo ou multimédia. La diversité de l'information est aujourd'hui si grande qu'il devient difficile d'en faire un panorama, une typologie exhaustive. N'importe quel type de document peut aujourd'hui être numérisé et publié sur Internet sous divers formats.

² Deja.com (<http://www.deja.com/usenet>) et Remarq.com (<http://www.remarq.com/>) permettent notamment une recherche sur l'ensemble des messages envoyés au travers d'usenet.

- **Au niveau de sa source** : Le web donne accès de manière indifférenciée à des informations validées par les organismes qui les diffusent (par exemple une organisation internationale ou un média traditionnel tel qu'un quotidien) et à des informations dont on ignore totalement la source, elles sont donc sujettes à tous les types de désinformation possibles (volontaire ou non).

- **Non-stable, versatile :**

Les informations publiées sur le web ne représentent jamais une entité finie comme peut l'être une monographie, elles n'ont pas de forme et de localisation définitives. Elles sont modifiées et évoluent en permanence. Il est donc impossible d'acquérir définitivement une information, il est uniquement possible de pointer sur elle, en espérant qu'elle ne disparaisse pas.

- **Répartie et non-contrôlée:**

Le web n'a pas de point central, pas d'index global recensant tous les serveurs ou toutes les informations. Il n'a pas non plus d'autorité de contrôle ou de censure.

2.1.2 Le père Noël n'existe pas

Pourquoi une information se trouverait-elle sur Internet ? Vous êtes-vous déjà posé cette question ? Malgré toute l'attention médiatique dont il dispose, il faut constater que le réseau Internet reste un concept flou pour la plus grande majorité, et pas uniquement pour les personnes mal à l'aise avec l'informatique et les nouvelles technologies. Même les professionnels de l'information ont parfois une image assez peu précise d'Internet. L'idée dominante est aujourd'hui la suivante : Internet est une mine d'informations. Certes, il apparaît difficile de ne pas être en accord avec cette idée. Mais l'image d'Internet, sa représentation collective, ne s'arrête pas là. Etant donné la richesse, je dirais presque l'opulence, de l'information disponible, les gens en arrivent à développer un véritable phantasme, à savoir : " quel que soit mon besoin d'information, je trouverai une réponse adéquate sur Internet ". Des idées comme celles-ci relèvent de l'irrationnel. Qu'est-ce qui soutient cette idée qu'Internet peut répondre à tous les besoins d'information ? Le réseau n'est pas un oracle, et c'est encore moins une bibliothèque géante alimentée par une armée de lutins bénévoles !

Il est capital de comprendre que si une information se trouve sur Internet, ce n'est pas pour répondre à notre besoin, mais *d'abord et avant tout* pour répondre au besoin de la personne qui l'a publié. Ce besoin pouvant être un besoin de promotion, de communication, de recherche, de collaboration, etc...

Cette notion – le pourquoi de la présence d'une information sur Internet - est essentielle, elle fait partie des concepts clé d'Internet, et elle est souvent ignorée. S'il est vrai qu'à l'origine du web, on trouve des chercheurs et des universitaires qui avaient pour but de faire partager gratuitement les fruits de leurs travaux et d'établir des collaborations avec leurs pairs; cette réalité a aujourd'hui été complètement dépassée par une autre logique : celle du business et du commerce. La présence des entreprises sur le web répond à cette logique marchande. Et tous les services "gratuits" que le web offre aujourd'hui, ne sont pas gratuits par philanthropie, mais sont financés par la publicité et la fréquentation des internautes. Le but des sociétés qui offrent ces services est d'atteindre la rentabilité. Pour autant, la volonté de partage du savoir ainsi que la collaboration gratuite et constructive n'ont pas disparu du web, elles n'en représentent simplement plus les aspects dominants.

Oui, le web tient ses promesses et il est bien un formidable outil de communication, mais, comme toutes choses ici bas il est soumis à un certain nombre de réalités

économiques et financières. Il faut donc, lorsque l'on fait des recherches sur Internet, savoir laisser de côté ses "fantasmes documentaires", et avoir à l'esprit qu'Internet est également un instrument commercial et pas seulement un "formidable puits d'informations et de connaissances".

2.1.3 Les outils de recherche

Ce chapitre constitue un panorama des outils et possibilités de recherche qu'offre le web. Pour plus de clarté, ceux ci sont divisés en plusieurs catégories.

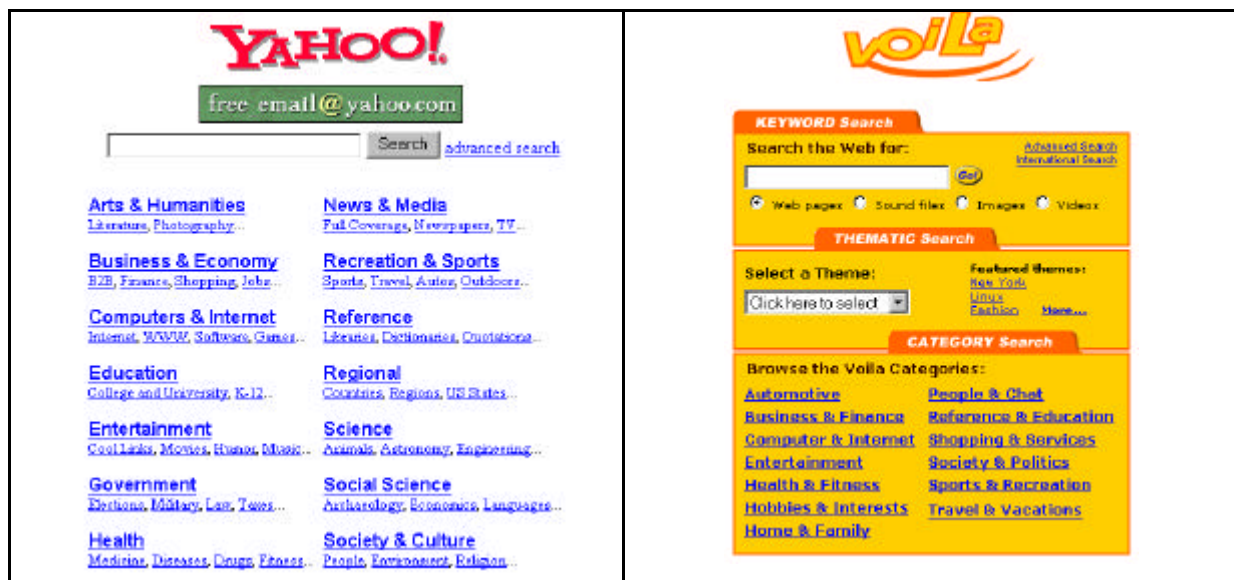
- annuaires
- moteurs de recherche
- meta-moteurs on-line
- agents intelligents
- outils spécialisés ou spécifiques

2.1.4 Annuaires

Historiquement, les annuaires ont été les premiers outils de recherche disponibles. Les annuaires ou répertoires sont de grandes listes de liens structurés d'une manière ou d'une autre, généralement par une arborescence ressemblant à celle d'un thésaurus.

Il est indispensable de bien faire la différence entre un moteur de recherche et un annuaire, car en pratique il est assez facile de confondre ces deux outils diamétralement opposés. Ils sont faciles à confondre parce qu'esthétiquement, ils se présentent souvent de la même manière.

Comparons par exemple deux interfaces, respectivement celle de "Yahoo!" et de "Voila".



Ces deux outils ne présentent que très peu de différences au point de vue de leur forme et de leur méthode de recherche : ils présentent tous deux les mêmes possibilités de recherche, à savoir : une recherche via une liste de catégories (qui sont assez similaires : business, computer & Internet, reference, sports etc..) ainsi qu'une recherche via un moteur et ceci grâce à un ou plusieurs mots-clés. Mais ces deux sites n'ont pourtant absolument rien en commun au niveau de leur contenu. Yahoo!, un annuaire, propose

une sélection de plusieurs dizaines de milliers de sites. Ces sites ont été sélectionnés, décrits et classés par des personnes et non pas de manière automatique, il y a donc une validation intellectuelle de leur contenu. Ces sites ont donc été retenus pour la qualité et l'intérêt des informations qu'ils proposent. Lorsque l'on effectue une recherche par mots-clés dans la base Yahoo!, celle-ci s'effectuera uniquement sur les titres et descriptions des sites sélectionnés.

Quant au contenu de Voilà, il est composé du texte intégral de plusieurs dizaines de millions de pages web. Ces pages ont été recensées de manière automatique et (en principe) sans filtrage de contenu. Les sites présentés dans les catégories proposées par l'interface de Voilà ont eux aussi été sélectionnés de manière automatique, grâce à une analyse sémantique de contenu. Le champ couvert par un moteur de recherche est donc énorme comparativement au contenu d'un annuaire.

Une requête effectuée sur un moteur de recherche tel que Voilà offrira donc un nombre beaucoup plus élevé de résultats que si elle avait été effectuée sur un annuaire comme Yahoo!. Cependant la pertinence et la qualité des résultats offerts par Yahoo! sera bien supérieure à celle de Voilà.

Les annuaires et les moteurs sont des outils très différents, ils ne sont pas conçus pour répondre aux mêmes besoins. Le caractère généraliste et la qualité de contenu des annuaires leur permettent de répondre rapidement à un grand nombre de questions courantes, mais ils sont très vite incapables de fournir des résultats face à des questions plus pointues. Peu importe que le contenu des annuaires soit bon si les informations que vous recherchez ne s'y trouvent pas !

Les annuaires représentent à l'heure actuelle, les seuls outils de recherche "humains", c'est-à-dire dont le contenu n'est pas généré automatiquement. Leur grande force est la valeur ajoutée de l'information proposée, celle-ci étant validée par un éditeur souvent spécialiste du domaine traité. Toutefois, il est nécessaire de nuancer cette notion de valeur ajoutée : la mission des personnes travaillant pour les annuaires étant de recenser assez rapidement un grand nombre de sites, ils ne consacrent pas toujours un temps suffisant à l'évaluation des sites qui formeront ensuite le contenu de l'annuaire.

Le plus grand et le plus célèbre des annuaires est bien entendu Yahoo!, qui recense plus d'un million de sites. Au-delà de Yahoo!, il existe des milliers d'annuaires généralistes et spécialisés, il serait d'ailleurs vain de tenter d'en établir une liste. Toutefois à titre indicatif, voici un aperçu de plus grands annuaires généralistes :

Nom	Adresse URL
Annuaire anglophones	
Britannica	http://www.britannica.com/
Dmoz – The Open Directory project	http://dmoz.org
Galaxy	http://galaxy.einet.net
Looksmart	http://www.looksmart.com
Lycos	http://www.lycos.com
Magellan	http://www.mckinley.com
NetGuide	http://www.netguide.com
Snap	http://www.snap.com
Yahoo!	http://www.yahoo.com
Annuaire francophones (français)	
Carrefour	http://www.carrefour.net

Francite	http://www.francite.com
Hachette	http://www.hachette.net
Nomade	http://www.nomade.fr
Yahoo! ! France	http://www.yahoo.fr/
Annuaire suisses romands	
Reference.ch	http://www.reference.ch
Search.ch	http://www.search.ch
Meta-annuaire ou site fédérateur	
Alpha Search	http://www.calvin.edu/library/searreso/Internet/as
The Argus Clearinghouse	http://www.clearinghouse.net/

Le dernier type d'annuaire présenté dans ce tableau, les meta-annuaire ou sites fédérateurs sont des outils extrêmement intéressants. Ils recensent les annuaires spécialisés, les sites faisant référence dans un domaine, ainsi que les sites "guides" offrant une collection de liens validés pour un thème précis. Ces meta-annuaire sont pour Internet ce que les bibliographies de bibliographies sont pour le monde documentaire traditionnel.

Trois tendances ou évolution dans le domaine des annuaires sont à relever:

1. La modification d'un statut d'annuaire à celui de portail

Il s'agit là d'une question de fréquentation et de politique de développement, de nombreux annuaires lorsqu'ils ont atteint une certaine masse critique de fréquentation, commencent à offrir des services supplémentaires (e-mail gratuit, news, achat et vente en ligne) et se transforment peu à peu de simples annuaires en véritables portails. (On se référera au chapitre 2.2 dédié aux portails pour plus d'information à ce sujet).

2. L'apparition d'un nombre de plus en plus important d'annuaires spécialisés

Le web ayant acquis un statut de véritable outil de travail, le nombre d'entreprises ou de groupes de travail créant leur propre annuaire a très rapidement augmenté. Les groupes d'intérêts et les communautés virtuelles (de personnes partageant les mêmes intérêts et actifs sur Internet) devenant plus structurées et mieux organisées, se sont également mis à produire des annuaires de qualité spécialisés dans leurs domaines. A titre d'exemple on citera Netserf³, un annuaire couvrant toutes les ressources ayant trait au moyen-âge.

3. L'apparition d'annuaires "open" ou ouverts

Le principe des annuaires open est le suivant : plutôt que de voir un seul individu ou une seule équipe assumer la responsabilité de construire et de valider l'ensemble du contenu d'un annuaire, ces tâches sont partagées. Les principales rubriques du répertoire sont confiées à la responsabilité de diverses personnes bénévoles, maîtrisant bien la rubrique ou la catégorie qu'elles désirent administrer. Ces personnes se chargent de rechercher et de décrire les sites présents dans la catégorie dont ils sont responsables. Cette nouvelle manière de gérer un annuaire permet de tirer parti des possibilités d'interaction et de

³ <http://www.netserf.org/>

synergie offerte par le web : un nombre très élevé de personnes peuvent ainsi travailler en collaboration sur le même annuaire, et ces personnes ont souvent une très bonne connaissance des ressources présentes sur Internet dans le domaine qui les concerne. Il aurait été tout à fait impossible de réunir un tel groupe de travail de manière "physique", c'est-à-dire dans un même lieu, une même institution.

Le dernier avantage des annuaires open est leur taille, ils sont mieux à même d'affronter la masse de document web en croissance constante, du fait du nombre de personnes travaillant ensemble. Si la coordination entre les divers contributeurs à un projet de ce type est bonne, l'outil en résultant peu être de très bonne qualité.

L'annuaire open le plus célèbre est "Dmoz – The Open Directory project".

2.1.5 Moteurs de recherche

Le terme "moteur" convient tout à fait pour nommer ces instruments de recherche, les plus célèbres du web. En effet, le contenu des moteurs est réuni de manière entièrement automatisée, il est donc dépourvu de toute subjectivité humaine. Leur principe est simple, ce sont des indexes géants du web. Lorsque l'on effectue une recherche sur un moteur, cela revient à consulter l'index d'un document de plusieurs millions de pages – le web.

Comment sont générés ces index, véritables bases de données titanesques ? Le principe est le suivant : les moteurs de recherche fonctionnent tous grâce à des applications appelées "robots" ou "spiders". Ces robots sont des agents automatiques qui parcourent le web en permanence, des espèces de "machines à surfer géantes" qui naviguent de page en page en suivant tous les liens qu'ils rencontrent, ils progressent ainsi de manière hiérarchique, un peu à la façon d'une arborescence qui se développe. Les robots récupèrent l'intégralité du vocabulaire pour chaque page qu'ils parcourent, et en utilise une partie pour constituer le contenu des index (pour les pages très longues seul le début de la page est pris en compte). Index sur lesquels l'utilisateur effectuera sa recherche.

Toute notion de tri selon des critères de qualité ou selon "l'émetteur" de l'information sont absents de ce processus. Cette opération est totalement "mécanique".

La génération automatique du contenu des moteurs de recherche en fait les outils les plus exhaustifs du web, ils sont donc essentiels à toute recherche efficace, mais leur taille en font des outils difficiles à manipuler.

Chaque moteur de recherche contient le texte intégral indexé de plusieurs millions de pages web, toutefois aucun d'entre eux n'est assez performant pour indexer la totalité des pages présentes sur le web.

La mise en place d'un moteur de recherche permettant l'indexation de million de pages étant un investissement extrêmement important, les acteurs sur ce marché ne sont pas légion. En voici une liste :

Nom	Adresse URL
Moteurs mondiaux	
Answers	http://www.answers.com.au/
AOL search	http://search.aol.com
Altavista	http://www.av.com
Excite	http://www.excite.com
Fast	http://www.alltheweb.com
Go	http://www.go.com
Google	http://www.google.com
Hotbot	http://www.hotbot.com
Infoseek	http://www.infoseek.com
Iwon	http://www.iwon.com
Lycos*	http://www.lycos.com
Magellan*	http://www.magellan.com
MSN	http://www.msn.com
Netscape	http://home.netscape.com/
Northern Light	http://www.northernlight.com
Snap*	http://www.snap.com
Voilà	http://www.voila.com
WebCrawler	http://www.webcrawler.com/
WebTop	http://www.webtop.com/
Moteurs francophones	
Ecila	http://www.ecila.fr
Excite France**	http://fr.excite.com/
Infoseek France**	http://www.infoseek.com/Home?pg=Home.HTML&sv=FR
Lokace	http://www.lokace.com/
Voilà	http://www.voila.fr
Moteurs suisses	
Sear.ch (Altavista suisse)**	http://www.sear.ch/

*Lycos, Magellan et Snap ont tous, dans un premier temps, été des annuaires, ils ont ensuite élargi leurs services et sont également des moteurs de recherche.

**Certains moteurs proposent des versions "locales" destinées à un pays en particulier. Il existe deux cas de figure concernant ces versions locales. Il peut s'agir de la même base de données, du même index mais auquel on applique des limitation de langues ou de nom de domaine (.fr ou .ch) ou alors d'un autre index réalisé spécifiquement pour la version locale.

Afin de mieux saisir qu'elle peut être la valeur des moteurs présentés dans le tableau précédant, voici un autre tableau présentant le champ couvert par un certain nombre de ceux-ci. On obtient une image de ce champ grâce à une estimation de la taille de la base de données de chaque moteur.

Moteur de recherche	Estimation de la taille de la base de données en millions de page
Google	560
Weptop	500
AltaVista	350
Fast	340
Northernlight	265
Excite	250
Inktomi	110
Go	50

Tableau extrait du site searchenginewatch.com⁴
(chiffres datés du 6 juin 2000)

Les deux principaux défis auxquels font actuellement face les moteurs de recherche sont les suivants :

- Comment faire face à l'augmentation exponentielle du nombre de documents sur le web ? Le plus grand moteur (Google⁵) ne couvre actuellement "que" 560 millions de pages soit un quart du web.
- Comment proposer des outils d'interrogation qui permettent d'obtenir rapidement des résultats pertinents dans une base de plusieurs millions de documents ?

Dans les deux cas, il s'agit pour les moteurs de recherche de trouver des solutions techniques et méthodologiques pour tenter de faire face à ces problèmes.

Il est aussi nécessaire, lorsque l'on parle de la taille des moteurs de recherche de parler du problème des liens morts (*dead links*). Les liens morts sont des sites ayant été déplacés ou supprimés mais dont l'adresse est toujours dans la base de données des moteurs de recherche. Suite à une interrogation sur un moteur, des liens morts se trouveront parfois dans la page de résultat et lorsque l'on clique sur ceux-ci pour accéder à la page désirée, un message d'erreur s'affichera pour nous signaler que la page n'existe plus.

La quantité de liens morts que contient un moteur de recherche est un bon critère pour en déterminer la qualité et la mise à jour.

Les options de recherche des différents moteurs

Pour des raisons de stratégie et de concurrence, l'hétérogénéité est plutôt la règle parmi les interfaces des moteurs de recherche, ils présentent chacun un nombre d'options et de critères de recherche qui leur sont propres. Les différentes interfaces sont également

⁴ <http://searchenginewatch.com/reports/sizes.html>

⁵ <http://www.google.com/>

plus ou moins conviviales, et nombre de moteurs proposent une "recherche avancée" permettant d'utiliser un plus grand nombre d'options.

Il ne s'agit pas ici de dresser un tableau comparatif de ces différentes options et syntaxes de recherche, mais seulement de présenter les plus répandues et les plus intéressantes d'entre-elles. Certains sites proposent déjà des tableaux de comparaison à jour et très complets⁶.

Limites de recherche

- Par date
Fenêtrage par dates extrêmes (du 1 janvier 2000 au 21 juillet 2000), plus récent que, pas plus récent que.
- Par domaine
.ch, .fr, .net, .com, .org, etc...
- Par serveur ou nom de domaine
unige.ch, unil.ch, geneva.ch
- Par langue
- Par titre de document
- Par medium (possibilité de rechercher des images, images en mouvement ou documents sonores)⁷

Recherche boudéenne

La très grande majorité des moteurs de recherche offre la possibilité d'effectuer des recherches boudéennes. Si aucun opérateur n'est spécifié dans l'équation de recherche, comme dans l'équation suivante:

The image shows a search engine interface with a yellow background. At the top, there are five tabs: 'Search' (highlighted), 'Advanced Search', 'Images', 'MP3/Audio', and 'Video'. Below the tabs is a search bar with the text 'charlie parker' and a 'Search' button. To the right of the search bar is a dropdown menu labeled 'any language'. Below the search bar, there is a tip: 'Tip: Use quotes for multi-word phrases.'

la recherche effectuée variera d'un moteur à l'autre, elle peut être interprétée comme:

- charlie AND parker
- charlie OR parker
- "charlie parker" en temps que phrase, c'est-à-dire comme une chaîne de caractères non dissociables

⁶ <http://www.abondance.com/outils/baromonde.html>

Comparatif réalisée par Olivier Andrieu sur son site "Abondance.com"

http://www.adbs.fr/adbs/sitespro/lardy/outil_ca.pdf

Tableau comparatif réalisée par Jean-Pierre Lardy, sur son site " La recherche d'information sur l'Internet"

<http://searchengineshowdown.com/features/byfeature.shtml>

Tableau comparatif réalisé par Greg Notess, sur son site " Search engine showdown".

⁷ Les outils de recherche permettant la recherche sur les images/images en mouvement et enregistrements sonores en sont encore au stade embryonnaire : ce type d'information pose notamment d'importants problèmes de droit d'auteur. A terme, il est possible que la résolution d'une partie de ces problèmes légaux (notamment par le biais de formats sécurisés, empêchant une utilisation illégale) amène une meilleure structuration de la diffusion de document audiovisuel sur le web et donc de meilleures possibilités de recherches.

Ces différences d'interprétation proviennent de la configuration par défaut de chaque moteur. Le " OR " comme opérateur par défaut augmente bien entendu le taux de bruit de manière significative.

Les opérateurs booléens généralement utilisés sont :

AND, +	Présence obligatoire du terme suivant l'opérateur
OR	Présence obligatoire de l'un ou l'autre des termes suivant et précédant l'opérateur
NOT, AND NOT, -	Absence obligatoire du terme suivant l'opérateur
NEAR	Présence obligatoire du terme suivant l'opérateur à proximité (à un nombre de caractères définis) du terme précédant l'opérateur

Troncature

La troncature est également une option proposée par de nombreux moteurs, mais à différents degrés. La troncature en début ou milieu de terme n'est notamment pas supportée par tous les moteurs.

Recherche par liens

La recherche par lien est une méthode particulièrement intéressante, c'est une sorte de recherche par citation. Cette recherche offre la possibilité de rechercher tous les sites ayant un lien hypertexte sur une page ou un site particulier.

Il est ainsi possible de reconstituer un réseau de sites et de personnes actives dans un domaine particulier. La démarche de recherche par lien est une forme de "surf-inversé" ou l'on ne visite pas les pages recommandées par un site, mais les pages ayant recommandé ce site.

Mots vides

Certains termes courants sont considérés comme vides : c'est-à-dire que les moteurs de recherche n'en tiennent pas compte et les éliminent simplement de l'équation de recherche. La liste de mots vides varie elle aussi d'un moteur à l'autre, les chiffres sont parfois considérés comme "mots vides". Les moteurs suivants recherchent tous les termes sans exception : AltaVista, Fast, Google, Infoseek, Lycos, Northern Light.

Tri des résultats par le biais de termes de tri

Altavista propose notamment cette option, qui fonctionne de la manière suivante : en plus de notre équation de recherche, il est possible de spécifier un ou plusieurs termes de tri. Les documents correspondant à notre équation seront affichés, mais parmi ces documents, ceux contenant également nos termes de tri figureront en tête de liste. Les termes de tri sont une sorte de critère de pertinence.

2.1.6 Meta-moteurs on-line

Les meta-moteurs on line sont des outils un peu particuliers : leur but est de permettre l'interrogation de multiples moteurs en une seule opération. Les meta-moteurs ne se présentent pas différemment des autres moteurs de recherche, si ce n'est que leur interface n'offre pratiquement pas d'options de recherche.

Concrètement les meta-moteurs on-line procèdent de la façon suivante : lorsqu'on leur soumet une équation de recherche, ils vont retransmettre instantanément cette équation

à toute une série de moteurs de recherche, attendre les réponses respectives de ceux-ci, éliminer les doublons, trier les résultats et en présenter une synthèse.

Cette opération est réalisée de manière tout à fait rapide, une interrogation sur un meta-moteurs étant très légèrement plus lente que sur un moteur classique. L'idée clé des meta-moteurs étant bien évidemment de gagner du temps, d'obtenir un résultat rapide et complet sans avoir effectué une multitude d'interrogation sur un grand nombre de moteurs.

L'idée est excellente, cependant les meta-moteurs ne sont pas la panacée de la recherche qu'ils prétendent être. Ils ne remplacent en aucun cas de multiples interrogations effectuées séparément sur chaque moteur, ceci pour les raisons suivantes:

- L'absence quasi-totale d'options de recherche, les meta-moteurs se limitant au mieux à des interrogations bouléennes simples. Il n'est pas réellement possible d'effectuer des recherches performantes dans ces conditions.
- Le nombre d'occurrences fournies par les meta-moteurs n'est jamais aussi élevé que le nombre obtenu en effectuant la même recherche sur chacun des moteurs. Les deux types de recherches ne sont donc PAS équivalentes.
- Deux des plus importants moteurs (Hotbot et Northernlight) ne sont pas interrogés par les meta-moteurs. Ces deux moteurs sont deux sources précieuses et représentent un champ important du web qui n'est pas exploitable via les meta-moteurs.
- L'hétérogénéité des types de sources interrogées, en effet, les outils interrogés par les meta-moteurs ne sont pas tous des moteurs de recherche, certains annuaires (notamment Yahoo!) sont parfois également interrogés. Cette tendance à mélanger plusieurs types d'outils peut être utile dans le cas d'une recherche ponctuelle ou succincte, mais n'est pas vraiment justifiable dans le cadre d'une recherche plus élaborée.
- Pour des raisons de rapidité de connexion, une recherche sur un meta-moteur effectuée plusieurs fois de manière consécutive ne fournira pas des résultats similaires. La raison de cette disparité est la subordination des meta-moteurs à d'autres moteurs. Lors d'une interrogation, si l'un des moteurs sollicités tarde à fournir une réponse, le meta-moteur se passera tout simplement des résultats de celui-ci. La stabilité des meta-moteurs est donc toute relative.

Il est également utile de signaler que le genre de service offert par les meta-moteurs n'est pas très apprécié par les véritables moteurs interrogés car les bannières publicitaires des moteurs (Altavista, Excite, Infoseek) ne seront pas visualisées par les utilisateurs alors que ces derniers bénéficieront tout de même des résultats.

Voici une liste des principaux meta-moteurs on-line.

Nom	Adresse URL
Meta-moteurs	
Cyber 411	http://www.cyber411.com
DogPile	http://www.dogpile.com
Inference Find	http://www.infind.com
MetaFind	http://www.metafind.com
Metacrawler	http://www.metacrawler.com
Profusion	http://www.profusion.com

SavvySearch	http://www.savvysearch.com

2.1.7 Agents intelligents

Les agents intelligents, contrairement aux autres outils de recherche présentés dans ce chapitre, ne sont pas des sites web ou des services accessibles en ligne, mais ce sont des logiciels. Ces logiciels s'installent sur un ordinateur client, et ils peuvent être paramétrés de différentes manières. Le qualificatif "intelligent" est utilisé pour refléter la capacité de ces logiciels à effectuer toute une série de tâches et d'actions de manière automatique. Les agents intelligents sont parfois également appelés "meta-moteurs off-line", cette seconde dénomination leur a été attribuée car la fonction principale des agents intelligents est similaire à celle des meta-moteurs, à savoir : interroger de multiples sources et moteurs de recherche on-line et présenter une synthèse des résultats obtenus. Afin d'éviter au chercheur d'avoir à effectuer ces opérations de manière manuelle. Cependant le terme "intelligent" serait quelque peu usurpé si les agents se limitaient simplement à interroger plusieurs moteurs de recherche. En effet, contrairement aux meta-moteurs, les agents intelligents sont des produits payants, ils ne sont pas uniquement financés par la publicité. Les services offerts par les agents intelligents sont donc d'un tout autre ordre que ceux proposés par les meta-moteurs.

Comment fonctionne un agent intelligent ?

Dans un premier temps, sur la base de notre équation de recherche, l'agent ira interroger une multitude de sources d'informations (annuaires, moteurs de recherche, bases de données spécialisées, newsgroups, sites spécialisés etc...). Puis dans un second temps, il se chargera de rapatrier ces résultats, c'est-à-dire qu'il copiera les pages web correspondantes aux résultats de la recherche sur votre ordinateur. Cette opération de copie des informations présente deux avantages :

- Une fois tous les résultats rapatriés, il est possible de les consulter sans être connecté à Internet.
- L'agent peut effectuer des tris et opérations complexes sur la base de ces résultats. Les doublons sont par exemple éliminés, et il si le tri proposé par l'agent ne convient pas, il est possible d'affiner la recherche grâce à des tris multiples.

Les possibilités offertes par les agents intelligents sont multiples :

- L'agent est configurable, il est possible de sélectionner quels moteurs de recherche et sources seront utilisés et ceci pour chaque recherche.
- Il est possible de programmer les agents afin d'effectuer la même recherche à intervalle régulier. L'agent devient alors un outil de veille permettant de suivre l'évolution d'un domaine particulier.
- Les agents interrogent un nombre de sources bien plus important que les meta-moteurs, et les sociétés qui créent les agents proposent très souvent des mises à jour, ajoutant aux agents la capacité d'interroger de nouveaux moteurs de recherche et de nouvelles sources.
- Les agents permettent d'effectuer des recherches par domaines. Les différentes sources utilisées par les agents sont classées. Ainsi si l'on désire effectuer une recherche ayant trait à l'actualité, l'agent n'interrogera pas la totalité de ses sources mais uniquement celles à même de fournir des réponses adaptées (par exemple des

bases de données de dépêches, les news et archives de divers quotidien, les grands sites d'informations etc.). Grâce à cette sélection de sources, les agents se transforment en outils de recherche spécialisés dans un domaine. A ce titre, ils sont qualifiés d'outils de recherche sectoriel. Il est évident que dans ce type de recherche sectorielle, leur pertinence est sans comparaison avec celle d'un simple moteur de recherche. Certains agents se sont spécialisés dans la recherche de produits au meilleur prix possible (livres, logiciels etc...), ceci en interrogeant notamment des catalogues de vente en ligne.

- Les méthodes de tri et d'exploitation des résultats est un autre atout des agents. Ces méthodes sont multiples, les plus intéressantes se basent sur une analyse sémantique des résultats. Il est ainsi possible d'indiquer à l'agent qu'un document trouvé est particulièrement pertinent, et l'agent s'efforcera (grâce à une analyse sémantique) de rechercher d'autres documents au contenu semblable.

Les agents intelligents sont des produits répondant à la nécessité d'effectuer des opérations de recherche d'information avancées et régulières (tel que la veille stratégique). Les agents intelligents sont encore appelés à évoluer durant les années à venir, notamment avec l'évolution des techniques d'analyse automatique de contenu qui leur permettront une plus grande autonomie et une plus grande "intelligence" dans la sélection de documents pertinents.

2.1.8 Outils spécifiques

Au-delà des outils traditionnels que sont les différents moteurs de recherche et autres annuaires, on trouve également sur Internet toute une série de possibilité recherche, qui sont ici réunis sous l'étiquette d'"outils spécifiques". Ces outils ne partagent pas de caractéristiques communes, si ce n'est celle d'être destinés à une utilisation plus ciblée que les autres moyens de recherche présentés au cours de ce chapitre.

Il existe bien entendu un très grand nombre d'outils répondant à des approches et à des besoins différents, notamment:

- Les pages jaunes, blanches et les répertoires d'adresses e-mail.
- Les outils permettant de trouver un newsgroup ou une liste de discussion spécifique. Certains outils permettent également de rechercher les messages et archives des newsgroups.
- Les répertoires de bases de données ou de moteurs de recherche sectoriels. Des milliers de bases de données de taille plus ou moins importante sont accessibles via le web, la difficulté étant de les localiser, puisque les informations contenues dans ces bases ne sont pas indexées par les moteurs de recherche. Ces bases de données représentent un type d'information que l'on qualifie parfois de "web invisible" (voir chapitre 2.1.12). Elles sont donc recensées par ces répertoires. Ceux-ci recensent également des moteurs de recherche spécifiques, c'est-à-dire dédiés à un secteur ou à un domaine précis. Ces répertoires remplissent parfois une fonction de meta-moteur pour les outils qu'ils recensent, c'est-à-dire qu'ils permettent d'interroger une multitude de sources en une seule opération.
- Les outils basés sur une approche géographique, ceux ci sont généralement développés pour les régions et pays ne comptant qu'un petit nombre de sites Internet (notamment l'Afrique).

Nom	Adresse URL
Pages jaunes et répertoire d'adresse e-mail	
Les pages jaunes suisses	http://www.pagesjaunes.ch
Infospace	http://www.infospace.com
Bigfoot	http://www.bigfoot.com
Newsgroup et listes de discussions	
Deja	http://www.deja.com/usenet
Liszt	http://www.liszt.com
Remarq	http://www.remarq.com/
Répertoires de base de données et de moteurs sectoriels	
All seeing Eye	http://www.streeteye.com/cgi-bin/allseeingeve.cgi
Invisibe Web	http://www.invisibleweb.com
The BigHub	http://www.thebighub.com
Webdata	http://www.webdata.com
Outils géographiques	
Africa south of the sahara : selected Internet resources	http://www-sul.stanford.edu/depts/ssrg/africa/guide.html

2.1.9 Trouver ce que l'on cherche ou la recherche d'une information pertinente

Effectivement, à la fin tout se résume à cette simple question : Comment trouver ce que l'on cherche ? Les outils actuels peuvent-ils nous fournir des réponses pertinentes et satisfaisantes ?

Plusieurs facteurs entrent en ligne de compte pour répondre à cette question :

- La taille gigantesque du web (2,1 milliards de page)
- La non-structuration de l'information
- La taille des moteurs de recherche :
Ces derniers sont loin de couvrir l'ensemble du web, mais ces moteurs recensent en moyenne plus de 100 millions de pages, ce qui, on l'imagine aisément, est bien suffisant pour "noyer" sous une masse de bruit énorme n'importe quelle information.
- La sous-utilisation des capacités des moteurs de recherche :
En effet, même si pratiquement tous les moteurs offrent des options de recherche avancées, la majorité des personnes se borne à effectuer des interrogations très simples. Ces interrogations sont à 90% des cas composées d'un seul ou de deux mots-clés. Ceci est tout à fait normal, la grande majorité des utilisateurs d'Internet n'ayant jamais été formée à effectuer des requêtes booléennes complexes ou ignorant simplement cette possibilité.

Et quel genre de résultats peu bien fournir un moteur de recherche comptant plus de 100 millions de références lorsque l'on effectue une requête simple, composée d'un seul terme ? Quelle chance la réponse du moteur a-t-elle d'être pertinente et de répondre aux attentes du chercheur ?

Tous ces facteurs et le constat que des recherches booléennes ne suffisaient plus à effectuer des recherches efficaces, ont poussé les gestionnaires des moteurs de recherche à trouver de nouvelles solutions afin de les rendre plus utilisables et d'en améliorer la pertinence. Quelques une de ces solutions seront détaillées dans le paragraphe qui suit.

Le tri des résultats : un chemin vers la pertinence

Le tri des résultats est un élément clé de la recherche, car dans bien des cas seuls les premiers résultats d'une recherche seront consultés par les utilisateurs (généralement les 10 ou 20 premiers sites). L'enjeu lors d'une recherche n'est donc pas de savoir qu'une page à été indexée dans la base de données, mais d'avoir la possibilité de "rappeler" cette page et ceci en tête de liste des résultats et non pas dans les 1000 premières réponses.

Chaque moteur a sa propre méthode de tri des résultats, celle-ci est souvent basée sur le nombre d'occurrence du terme recherché dans le texte des documents, dans les meta-tags ou dans le titre.

On ne connaît pas précisément les critères de tri de moteurs de recherche, ils n'en parlent pas, et ceci pour une bonne raison : éviter que les sociétés qui publient sur le web n'abusent de la situation et ne cherchent à se placer au sommet des résultats lorsqu'un certain terme est recherché (par exemple faire apparaître son site en tête de liste alors que c'est le nom d'un concurrent qui à été recherché.)

Ces solutions classiques de tri des résultats sont maintenant épaulées par des approches différentes.



Narrow Your Search with Custom Search Folders™

Your search returned 9,620,368 items which we have organized into the following [Custom Search Folders](#):

-  [Search Current News](#)
-  [Special Collection documents](#)
-  [Web search engines & directories](#)
-  [Sports](#)
-  [World Wide Web](#)
-  [Canada](#)
-  [Web search utilities](#)
-  [Academic publishers](#)
-  [Microsoft Windows operating systems](#)
-  [Hotbot](#)
-  [Cars](#)
-  [History of architecture](#)
-  [Farms & farming](#)
-  [all others...](#)

US Patent 5,924,090

Illustration d'une catégorisation des résultats, pour une recherche sur l'équation : "web search"

Le moteur Google⁸ a introduit une nouvelle approche du tri des résultats. Le critère qui déterminera la place de telle ou telle page web au sein des résultats de recherche est la popularité de celle-ci. L'idée qui soutient cette démarche étant que si une page est très largement citée (le nombre de liens hypertextes pointant sur elle est élevé), celle-ci d'autant mieux classée. Cette méthode de tri permet d'assurer une (relative) qualité des résultats de recherche, mais condamne sans appel les sources moins connues ou moins populaires.

2.1.10 Tendances et perspectives d'avenir

La tendance actuelle, au niveau des sites offrant des moyens de recherche, est la multiplication des services : de plus en plus d'acteurs sur le marché de la recherche tentent de se positionner en tant que "moyen de recherche complet et unique". D'une part, les moteurs de recherche se sont généralement dotés d'annuaires ou d'accès thématique, d'autre part, des sites qui jusque là n'étaient que des annuaires proposent également un moteur de recherche.

De plus, tous les outils de recherche populaires tentent d'augmenter et d'exploiter leur fréquentation en proposant des services de type portail. En résumé, on assiste, dans le domaine des annuaires et les moteurs de recherche (qui représentent la première génération d'outil de recherche), à une uniformisation des services proposés. Certains

⁸ <http://www.google.com/>

moteurs tentent toutefois d'innover en proposant des fonctions proches de la veille : Northern Light propose notamment la possibilité de sauvegarder une équation de recherche, d'effectuer cette recherche à intervalles réguliers et d'en communiquer les résultats par e-mail.

L'avenir de la recherche sur Internet se trouve sans aucun doute dans l'interrogation en langage naturel. Les premiers outils offrant une possibilité de recherche par le biais de phrases simples (et non de mots-clés) ont d'ailleurs commencé à faire leur apparition. Le site Ask Jeeves⁹ offre notamment cette possibilité, toutefois nous sommes encore très loin d'une véritable "compréhension" de la question par une intelligence artificielle. Ask Jeeves fonctionne grâce à un système de questions types auxquelles correspondent des réponses se trouvant dans des sources sélectionnées. Oingo¹⁰, un outil assez récent, offre une réelle possibilité d'effectuer des interrogations basées sur le sens et les concepts plutôt que sur la simple présence ou non d'un terme précis dans un texte. Si l'utilisateur utilise un terme d'interrogation ambigu, (par exemple java, qui peut être utilisé pour désigner l'île de java, du café ou encore le langage de programmation). Oingo proposera à l'utilisateur de choisir entre les différents sens du mot, pour fournir une réponse pertinente.

Il est certain que l'application des recherches en intelligence artificielle permettra à terme d'obtenir des outils plus évolués. Avec la mise en pratique de l'intelligence artificielle, on assistera à la transformation de moteurs de recherche, ils vont évoluer d'un état de singe savant vers celui de véritable machine à raisonner, capable d'effectuer des raisonnements logiques lorsqu'on leur soumettra une question. Le but recherché étant d'obtenir des moteurs capables de comprendre le contexte d'une interrogation en langage naturel, et de fournir une réponse adéquate.

Cette forme "idéale" d'interrogation est toutefois une perspective très lointaine. Dans un avenir plus proche on verra certainement apparaître de nouvelles méthodes de tri des résultats permettant une meilleure pertinence ainsi qu'une simplification des interfaces d'interrogation (le moteur Hotbot¹¹ faisant figure de pionnier, en rendant son interrogation de plus en plus simple et explicite, toutes les options de recherche étant accessibles via des listes de choix).

Les agents intelligents auront également une place de choix dans la panoplie du chercheur de demain, ils bénéficieront eux aussi des avancements en matière d'analyse automatique de contenu et d'intelligence artificielle.

Un autre service déjà offert par quelques agents se popularisera très certainement : il s'agit de la capacité de "trouver des documents similaires", en indiquant à un moteur de recherche qu'un document nous convient, il analysera celui-ci afin de fournir d'autres documents au contenu similaire. Deux outils offrent ce genre de service : le moteur de recherche webtop¹² auquel il est possible de soumettre un texte entier ou quelques paragraphes, webtop utilise ensuite ce texte comme critère de recherche. Le second outil est Nano¹³, un logiciel qui effectue à la demande une recherche sur le web après avoir analysé un document qu'on lui aura soumis (e-mail, fichier word, page web).

La recherche sur Internet recèle de nombreux domaines encore peu exploités (notamment la recherche sur les fichiers audiovisuels). Ces domaines vont très certainement évoluer dans les années à venir et les outils existants vont continuer à se transformer de manière rapide; il est assez hasardeux d'essayer à prédire quels seront les moyens de recherche de demain, tant le domaine évolue rapidement. Il est certain que les moteurs de recherche sont là pour durer, mais seront-ils capables de faire face à l'évolution exponentielle du web ? Une chose est certaine : ils

⁹ <http://www.ask.com>

¹⁰ <http://www.oingo.com>

¹¹ <http://www.hotbot.com>

¹² <http://www.webtop.com>

¹³ <http://www.nano.com>

ne dépasseront pas un certain stade de complexité dans la recherche. Ces moteurs étant financés par la fréquentation et la publicité, leur but premier n'est pas réellement de fournir le moyen de recherche le plus élaboré et le plus performant possible, mais d'atteindre un seuil de rentabilité. Au-delà d'un certain niveau de complexité, l'outil n'est plus rentable. Il s'agit avant tout pour les moteurs de faire face à leur concurrence. Le jeu de la concurrence est le vecteur principal qui pousse les moteurs à évoluer et à innover. Les approches réellement innovatrices proviendront donc peut-être d'outils qui ne sont pas des services financés par la publicité comme les moteurs de recherche, mais qui sont de véritables produits (tel que les agents intelligents).

2.1.11 Les stratégies de recherche

La première étape d'une recherche d'information est toujours la même : Identifier l'information que l'on recherche afin de déterminer l'outil le plus à même de fournir une réponse adéquate.

Dans ce cadre, il est indispensable de se demander si l'information que l'on cherche n'est pas accessible de manière plus simple et plus rapide dans une autre source qu'Internet.

Si ce n'est pas le cas, il s'agit de sélectionner l'outil à utiliser, la démarche consiste à choisir l'outil le plus spécifique possible :

Dans le cas d'une recherche d'adresse e-mail par exemple, on utilisera d'abord les répertoires d'adresses e-mail avant d'avoir recours aux moteurs de recherche. Cet exemple est une lapalissade, mais il illustre bien la méthode. Le fait de systématiquement commencer une recherche par un moteur n'est pas une bonne chose, cela risque de nous faire perdre du temps, et il n'est même pas certain que le moteur de recherche puisse nous fournir une réponse.

Certains critères de sélection de l'outil sont évidents et d'autres le sont un peu moins, par exemple dans le cas d'un problème informatique, d'un logiciel qui ne fonctionne pas correctement, la démarche qui consiste à aller du spécifique au générique sera la suivante :

- Vérifier si la solution ne se trouve pas sur le site de l'éditeur du logiciel
- Rechercher s'il existe un newsgroup susceptible de répondre à notre question, en utilisant le site <http://www.deja.com>. Si celui-ci existe, la solution de notre problème se trouvera certainement dans les archives de ce newsgroup.
- Si les deux premières possibilités ont échoué, c'est alors seulement qu'on se tournera alors vers un moteur de recherche.

Au-delà de la sélection de l'outil de recherche, la connaissance préalable de l'environnement d'une question est un atout majeur dans la recherche. Elle sera beaucoup plus rapide et aisée si le chercheur a une bonne connaissance des ressources web dans le domaine l'intéressant.

Dans le cas d'une recherche documentaire suivie (non ponctuelle) qui pourrait s'apparenter à une activité de veille stratégique, l'identification des sources est un processus assez lent. Il faut plusieurs semaines pour se faire une idée, une représentation précise des sites traitant du sujet nous intéressant et offrant un contenu digne d'intérêt et mis à jour. La qualité d'un site dépend en partie de sa mise à jour, et celle-ci ne peut être vérifiée qu'à long terme.

Utilisation des moteurs de recherche

Utiliser plusieurs moteurs

C'est une chose acquise; aucun moteur n'indexe l'intégralité du web, il est donc logique d'utiliser plusieurs moteurs différents afin d'élargir au maximum le champ de la recherche.

Il faut aussi avoir à l'esprit le fait que les moteurs indexent des parties *différentes* du web, et que le taux de recoupement entre les moteurs est relativement bas. Pour illustrer ceci, on pourra citer un test réalisé par Greg Notess¹⁴ selon le principe suivant :

Il effectue cinq recherches similaires sur quatorze moteurs de recherches différents, puis il compare les résultats afin de déterminer si les différents moteurs ont trouvé les mêmes pages. Seul 24% des pages ont été trouvées par plus de trois moteurs de recherche. Ce qui signifie que plus de trois quarts des informations ne peuvent être trouvée que par deux moteurs sur les quatorze utilisés pour cette recherche.

Ce test met en évidence la complémentarité des moteurs de recherche, il n'est dès lors plus possible de considérer une information comme inexistante sur le web sans avoir au maximum diversifié les outils de recherche.

Ce taux de recoupement peu élevé entre moteurs amène un autre constat : il est souvent beaucoup plus efficace de changer de moteur de recherche plutôt que de s'acharner à reformuler son équation de recherche sur un seul moteur.

Utiliser les plus "grands" moteurs de recherche – ceux recensant le plus grand nombre de pages

Pourquoi est-il important d'utiliser les moteurs de recherche recensant le plus grand nombre de pages ? On pourrait se dire qu'un grand réservoir de sites n'en contiendra que plus de bruit et décider de se limiter à des annuaires (Yahoo!) ou à de petits moteurs de recherche. L'utilisation des "grands moteurs" est capitale car ceux-ci sont la seule solution pour trouver des réponses à des questions précises qui sortent des domaines courants couverts par les annuaires.

Elargir la recherche

Face à une recherche infructueuse, l'élargissement de la recherche est une stratégie qui peut s'avérer payante :

Ne pas utiliser d'équation booléenne trop complexe, de ne pas utiliser de limite de date ou de langage. En fait, prendre le contre-pied des habitudes de recherche que l'on peut avoir en tant que professionnel de l'information.

Le fait d'utiliser une équation booléenne complexe et d'autres critères de recherche discriminants est une approche valable lorsque l'on a à faire à de l'information structurée comme dans une base de données bibliographique, mais l'absence de structuration dans le cas du web rend les critères de recherche précis parfois inefficaces.

Il est donc quelquefois justifié de tenter des recherches moins précises et moins complexes (ou d'utiliser le "OR" booléen), ceci augmentera bien sûr le bruit mais permettra peut-être d'éviter de laisser échapper une information.

¹⁴ <http://www.notess.com/search/stats/overlap.shtml>

2.1.12 Les limites de la recherche : Le web invisible

Les moteurs de recherche et les annuaires indexent et recensent des millions de pages, toutefois, il faut se rappeler que même réunis, ils sont très loin d'indexer l'intégralité du web. Ces pages laissées de côté par les moteurs de recherche et les annuaires constituent la première partie de ce qu'on appelle *le web invisible*. La seconde partie du web invisible est quant à elle composée d'informations qui pour des raisons techniques, ne pourront jamais être recensées par un moteur de recherche. On appelle aussi ces informations des *inindexables* (par opposition aux autres informations restant *inindexées*), ce sont :

- Le contenu des fichiers PDF (et autre format propriétaires).
- Le contenu des sites payant ou nécessitant un mot de passe.
- Les bases de données (par exemple les archives d'un quotidien), qui ne sont pas des pages HTML, et que les robots ne peuvent recenser.
- Certaines pages dynamiques utilisant des CGI, qui ne peuvent être indexées car elles sont créées à la demande par l'utilisateur.
- Les sites n'ayant aucun lien pointant sur eux (les robots des moteurs de recherche n'ont aucun moyen de les atteindre et donc ne peuvent les indexer, ces pages sont véritablement cachées).
- Certaines pages utilisant des frames.
- Les pages marquées par un code signalant aux robots des moteurs de recherche de ne pas les indexer.

L'invisibilité de certaines bases de données accessibles sur le web est parfois relative, car ces bases sont recensées par des répertoires spécifiques. Quelques uns de ces répertoires sont cités dans la section "Outils spécifiques" de ce chapitre.

2.2 Les portails

On entend aujourd'hui beaucoup parler de site portail, de portail (portal en anglais) ou parfois (terme un peu plus exotique) de "vortal". Que recourent toutes ces notions ? Qu'est-ce véritablement qu'un site portail ?

Pour répondre très simplement et très brièvement : le portail peut être qualifié de "porte d'entrée", de point d'accès par lequel l'utilisateur commencera son utilisation d'Internet.

2.2.1 La preuve vivante : Yahoo!

Historiquement le premier portail d'Internet fut Yahoo!, le célèbre annuaire. C'est véritablement avec Yahoo! qu'est née cette notion. Pour bien comprendre ce qui se cache derrière la genèse du concept de portail, il est intéressant de se pencher sur l'évolution qu'a suivi Yahoo! :

Au départ le site se limitait à un simple répertoire de liens, une sélection de sites. Cette sélection représentait l'unique contenu, le seul service offert aux internautes. On avait donc ici affaire à un carrefour qui dirigeait les internautes vers des sources ou sites susceptibles de présenter un intérêt.

Ce rôle d'aiguilleur du web a apporté à Yahoo! un grand nombre d'utilisateurs et c'est grâce à ce rôle qu'il est devenu la grosse société que l'on connaît aujourd'hui. Cette tâche de "redirection", de sélection des sources est le premier des éléments fondateurs du portail.

Au-delà de son succès "populaire", Yahoo! a créé une nouvelle habitude d'utilisation : le service de sélection des sites était si intéressant qu'il devenait un passage quasi obligé, ou pour présenter les choses autrement, si une personne était à la recherche d'information sur le web, son point de départ était quasi systématiquement Yahoo!. C'est cette utilisation qui est intéressante, l'idée qu'à chaque passage sur le web, une page soit systématiquement consultée.

Ce concept de fidélisation, de point de départ préférentiel est la deuxième idée clé du portail. Le portail se doit d'offrir un contenu qui soit si intéressant qu'il en devienne indispensable.

Si l'on regarde aujourd'hui à quoi ressemble la première page de Yahoo!. On y trouve bien plus qu'un répertoire, on notera pêle-mêle : l'accès à un e-mail gratuit, des actualités, tous types d'informations (économiques, boursières, culturelles, sportives etc...), des offres d'achat et de vente en ligne etc...

Yahoo! (et les autres portails) en sont bien entendu arrivés à proposer ces divers services et informations pour fidéliser les utilisateurs ainsi que pour tenter de conserver une position de "porte d'entrée généraliste".

2.2.2 Les différents type de portail

■ Les portails généralistes

Martin White dans son article "Portals : here today, gone tomorrow ? "¹⁵ défini les portails de la façon suivante :

¹⁵ WHITE, Martin, **Portals: here today, gone tomorrow ?**,

IN: Online information 99 : proceedings [of the] 23rd International Online Information Meeting, National Hall and Olympia 2, London, UK, 7-9 December 1999, Oxford, Learned Information Europe Ltd, 1999, p. 9-11

" [...] a site that generates revenues from advertising and sponsorship, and from shared revenues deals, through attracting high volumes of traffic ".

Cette définition du portail par son financement, si elle ne dit rien sur les activités du site, a au moins le mérite de correspondre à la réalité. Les portails généralistes se doivent d'être gratuits et de générer leur revenu grâce au trafic qu'ils engendrent (par le biais de bandeaux publicitaires et de liens directs avec des partenaires de vente en ligne).

Les portails généralistes sont souvent issus de fusion ou de collaboration entre des prestataires de contenu (médias, industrie du disque et du cinéma etc...) et des sociétés bien établies dans le monde d'Internet et des réseaux.

Du fait de l'avance considérable des Etats-Unis dans l'utilisation d'Internet, les grands acteurs dans le domaine des portails généraliste sont bien évidemment tous américains (Yahoo!, MSN et AOL étant les trois plus grands).

En Europe les portails généralistes sont fréquemment issus de la presse écrite, mais les géants américains tentent depuis quelque temps de développer des versions "locales" de leurs portails pour les différents pays européens (Yahoo!, est présent dans tous les grands pays d'Europe, Excite et AOL le sont également).

■ Les " Vortals ", portails verticaux ou portails spécialisés

S'ils répondent aux mêmes caractéristiques que les portails généralistes, ces sites sont spécialisés dans un domaine précis, il s'agit le plus souvent de spécialisation dans un domaine professionnel ou commercial (par exemple salesguy.com¹⁶, spécialisé dans la vente ou bakeryonline.com¹⁷, adressé – bien entendu - aux professionnels de la boulangerie). Mais on trouve aussi des portails verticaux non professionnels dans des domaines tels que le sport¹⁸.

Les " vortals " sont un type de site en pleine expansion actuellement. Le nerf de la guerre, la raison qui explique leur croissance importante à l'heure actuelle, est le commerce électronique. Les portails verticaux étant fréquentés par des publics très spécialisés, ils offrent une plate forme idéale pour le commerce électronique.

■ Les portails d'entreprise

Le portail d'entreprise ou EIP " Enterprise Information Portal " représente un type de portail extrêmement intéressant. Financé par son entreprise, le but du portail n'est plus d'être viable " commercialement ", mais il se doit d'être un outil complet d'information, de communication et de documentation. L'idée du portail d'entreprise est née de deux besoins :

- une meilleure exploitation - *par tous les collaborateurs de l'entreprise* - des ressources informationnelles et documentaires existantes au sein de l'entreprise.
- un besoin de structuration de l'information externe nécessaire à l'activité de l'entreprise.

La principale caractéristique du portail d'entreprise est de mélanger de façon transparente et organisée l'information interne et l'information externe, puis de présenter cette information de manière synthétique. A terme on peut imaginer

¹⁶ <http://www.salesguy.com>

¹⁷ <http://www.bakeryonline.com/>

¹⁸ <http://www.sportal.com> ou <http://www.mountainzone.com> pour les sports de montagne)

que le portail d'entreprise deviendra l'outil central d'une gestion globale de l'information et des ressources documentaires présentes dans une entreprise.

Le portail d'entreprise est une branche en pleine croissance, qui selon une estimation du Delphi Group devrait représenter un marché de 15 milliards de dollars d'ici 2003. Les grandes entreprises internationales devraient toutes être dotées de leur propre portail d'ici là.

Portails généralistes et portails d'entreprises : un tableau comparatif

	Portails généralistes et commerciaux	Portails d'entreprise
Quel est le support ?	Internet	Intranet, Extranet
Comment est-ce financé ?	Par la publicité	Par l'entreprise
Qui l'utilise ?	Le grand public	Les collaborateurs de l'entreprise
Quel type d'information y trouve-t-on ?	Tout type d'information	De l'information relative à l'entreprise et à ses activités

2.2.3 Les différentes fonctions d'un portail

■ La sélection d'information

La sélection d'information reste la première fonction - la raison d'être - du portail. Les portails sont nés en réaction à la surabondance d'information sur le web, pour donner des points de repère, pour orienter. Cette fonction d'orientation est réalisée en sélectionnant des informations, c'est-à-dire en pointant sur des informations et sites (payants ou gratuits) de qualité et d'intérêt. Le rôle des portails, grâce à cette sélection, est donc de fournir de l'information de type secondaire – de l'information sur l'information. Mais les portails peuvent également offrir de l'information primaire gratuitement, grâce à des partenaires possédant cette information. Pour les portails généralistes cela peut être des extraits de film ou de morceaux musicaux et pour un portail vertical ou spécialisé, cette information primaire pourra être des extraits d'articles ou de bases de données.

■ La fonction d'alerte

Les portails ayant pour vocation d'être fréquemment consultés, ils publient des " news ", remplissent une fonction d'alerte. En effet, la fréquence d'utilisation et le succès d'un site est proportionnel à la qualité et la fraîcheur de l'information que celui-ci fournit. Il s'agit quasiment d'un axiome dans le monde du web : seule une mise à jour constante et rapide saura instaurer une habitude d'utilisation. C'est dans cette optique, d'habituer l'utilisateur, de lui faire multiplier ses visites que la fonction d'alerte apparaît donc comme capitale pour les portails. Et ceci aussi bien pour les portails d'entreprise ayant pour but d'informer au mieux au sein de l'entreprise, que pour les portails commerciaux dont la survie dépend de la fréquentation.

Cette fonction d'alerte peut prendre diverses formes et différents niveaux de spécialisation selon les portails, elle peut aller des simples actualités (de type presse, journal télévisé) jusqu'à la publication d'informations très pointues, qui seraient le résultat d'une activité de veille stratégique dans un domaine spécifique.

■ La personnalisation de l'information

La plupart des portails proposent de l'information personnalisée. L'idée de base est la suivante : générer une version " personnalisée " du portail pour chaque utilisateur. L'internaute sélectionne un ou plusieurs thèmes selon ses intérêts, puis à sa prochaine visite, l'utilisateur ne trouvera pas la page " standard " du portail, mais une page personnalisée avec un accent particulier mis sur les informations relatives aux thèmes qu'il aura choisis. Cette pratique de personnalisation s'apparente au concept de la technologie PUSH.

■ Les services complémentaires

Les services complémentaires sont tous ces outils et informations proposés par les portails en plus de leur service principal à savoir la fourniture d'information. Ces services peuvent être extrêmement variés, à savoir : e-mail gratuit, forum, vente et achat en ligne, information météo, agenda, outils de gestion, calendrier ou encore menu de la cafétéria pour un portail d'entreprise, bref on peut imaginer une foule de services différents. Bien entendu, ces services sont adaptés aux publics spécifiques des différents portails.

2.3 Création de site web

Que signifie – en pratique - créer un site web à l'heure actuelle ? Quelles implications techniques se cachent derrière la réalisation d'un site web ? Quels objectifs souhaite-t-on atteindre lorsque l'on décide d'utiliser les technologies et le réseau Internet ? Ces questions seront traitées tout au long de ce chapitre.

2.3.1 Le web : communication et interaction

Le but premier d'Internet n'est-il pas depuis sa création (si l'on fait abstraction des objectifs un peu plus militaires d'Arpanet, l'ancêtre du réseau actuel) de permettre et d'améliorer la communication entre les différents utilisateurs du réseau ? Oui, la communication était bien le but recherché lors de création et tout au long des diverses évolutions qu'a suivi Internet. Parallèlement à cet objectif de communication, qui était déjà rempli dès la création d'Internet, le réseau a énormément évolué techniquement. Plusieurs facteurs ont rendu possible la situation dans laquelle nous nous trouvons aujourd'hui :

- La croissance des capacités de transfert, de la bande passante et des réseaux.
- L'augmentation de la puissance - et donc des possibilités – de la micro-informatique et des ordinateurs personnels.
- La croissance exponentielle du web dès 1994 et dans les années qui ont suivies, ce qui l'a amené à être de facto le standard mondial qu'il est à l'heure actuelle.
- Le développement des nouvelles techniques et possibilités liées au web, (les divers standards, langages de programmation et solutions techniques).

Toutes ces évolutions et ces développements ont fait d'Internet ce qu'il est aujourd'hui, à savoir qu'il n'est plus uniquement un moyen de communication mais un véritable outil de travail. Les développements techniques ont amené de nouvelles possibilités et à l'objectif premier de communication, s'est ainsi greffé un second objectif : l'interaction.

L'illustration idéale de cette évolution d'un but de communication vers un but d'interaction est le commerce électronique. Les technologies du web permettent à l'heure actuelle d'effectuer toutes les opérations nécessaires à une transaction commerciale (inscription en ligne, choix des articles dans un catalogue mis à jour en temps réel, paiement sécurisé).

Si auparavant la seule fonction d'Internet était l'échange d'information, les moyens techniques nous permettent maintenant d'interagir, de collaborer sur ces mêmes informations. A partir de là, un nombre énorme de développements et d'évolutions sont imaginables. Tous ces éléments nous amènent à la conclusion suivante : actuellement, la création d'un site ne se définit plus en termes de vitrine, de publication d'information, mais la technique offre aujourd'hui la possibilité de créer une véritable application aux fonctionnalités et services multiples.

2.3.2 Les technologies du web

HTML, la norme du web

Avec la popularité grandissante du web, la notion de codage des documents s'est assez largement démocratisée. La plupart des personnes utilisant le web savent plus ou moins précisément, ce qu'est une page HTML¹⁹, ou en tout cas en comprennent le principe. Les éditeurs WYSIWYG²⁰ (tels Dreamweaver ou Frontpage) ont également contribué à simplifier et démystifier les normes utilisées sur le web. A l'heure actuelle, la technique ne constitue plus véritablement un obstacle à la publication d'un document sur Internet. En effet, au niveau technique, le pas à franchir entre créer un document Word et publier un document simple sur le web n'est aujourd'hui plus très grand.

Parallèlement à cette démocratisation au sein du grand public de concept tel que l'HTML, il s'est opéré ces 3 dernières années, un abandon graduel de l'HTML au niveau de l'utilisation professionnelle.

C'est une réalité : à l'heure actuelle une grande majorité des sites de taille importante, des sites présentant des fonctions interactives ou nécessitant une mise à jour extrêmement fréquente ne sont plus gérés en utilisant des pages HTML.

Il est important de comprendre que le résultat final, la page que l'on obtient sur notre navigateur est toujours en HTML, c'est la genèse de cette page qui n'a plus rien à voir avec l'HTML.

Par analogie, on pourrait comparer le langage HTML au papier. HTML est en quelque sorte devenu la feuille A4 du web, et les applications et techniques qui permettent de gérer un site web, jouent le même rôle que le logiciel Word par rapport à cette feuille de papier. Il n'est pas possible, lorsque l'on regarde un document sur papier de déterminer grâce à quel logiciel il a été écrit, mis en page, etc... Il en va de même pour les sites web et l'HTML : notre page s'affiche à l'écran en HTML, mais il est difficile de déterminer comment cette page a réellement été générée, quel moyen technique se cache derrière sa création. Les pages générées de cette façon sont qualifiées de "pages dynamiques" par opposition à une simple page HTML, qui s'affichera toujours de la même façon, et qualifiée pour cette raison de "page statique". Ce sont les pages dynamiques qui sont à la base de toutes les possibilités d'interaction qu'offre le web. Ces techniques ont rendu possible un bon nombre des développements que connaît aujourd'hui Internet.

2.3.3 Pages statiques et pages dynamiques

Pour comprendre ce qu'est une page dite dynamique, il faut la comparer à une page HTML "classique". HTML n'est pas un langage de programmation, c'est une norme de mise en forme, de balisage des documents. Une page HTML contient un texte ainsi que des informations qui déterminent la façon dont ce texte sera affiché. On peut ainsi formater ce texte comme bon nous semble (mettre des sections de texte en gras, insérer des images etc...), mais il est impossible d'effectuer des opérations complexes au travers de l'HTML ou de réaliser un traitement différent de la page suivant les conditions (par exemple afficher la date du jour sur une page, ou afficher cette page en français plutôt qu'en anglais lorsqu'elle est consultée par un internaute utilisant un navigateur en français).

Une page HTML, à la manière d'un texte mis en page sur papier, sera donc toujours affichée de la même façon, c'est pour cette raison qu'elle est qualifiée de "statique". A l'inverse, une page dynamique, pourra être conçue pour effectuer quasiment n'importe quelle opération (gérer un site multilingue, accéder à une base de données, mettre à jour des informations en temps réel etc...). Toutes ces opérations sont réalisées au travers d'un langage de programmation. Les pages d'un site ne sont alors plus "écrites" en HTML, mais dans le langage de programmation choisi.

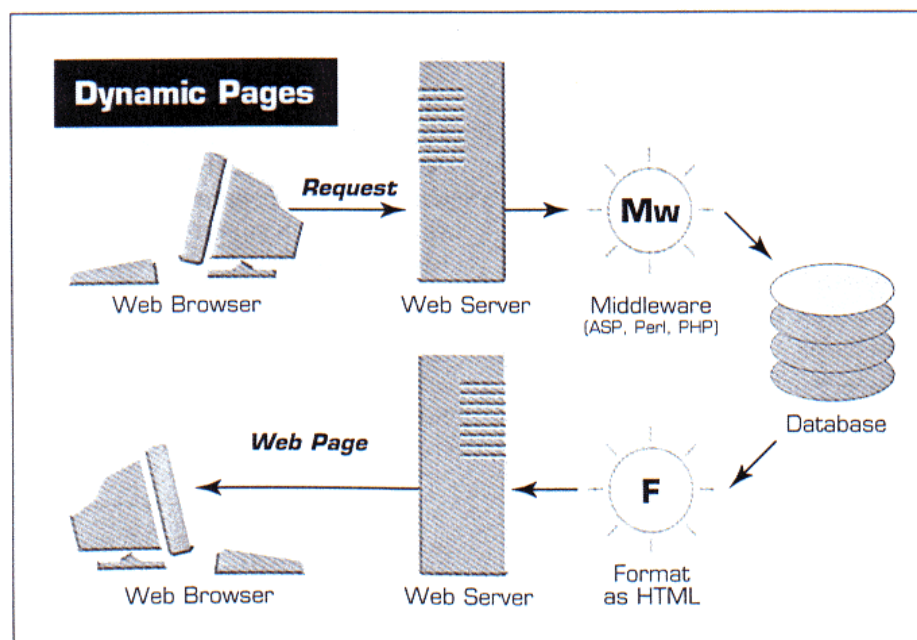
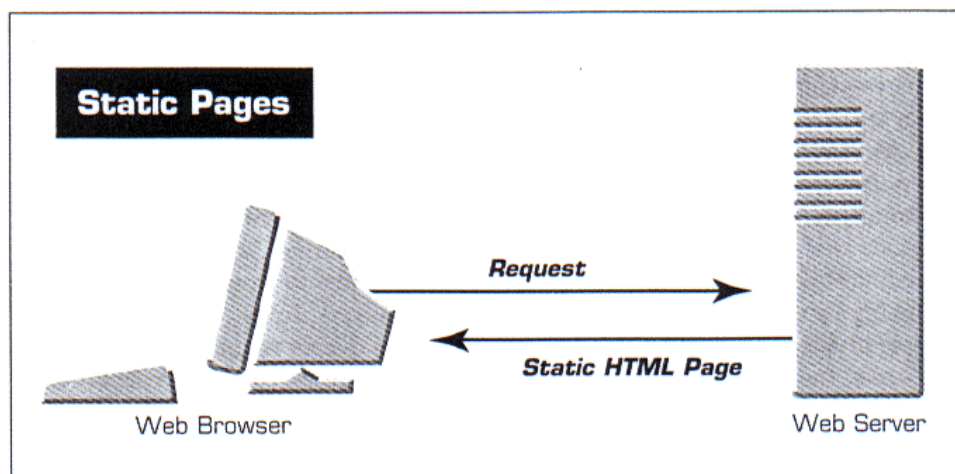
Cependant, les différents navigateurs (Internet Explorer, Netscape) sont uniquement capables d'afficher des pages écrites en HTML, alors comment les pages dynamiques, contenant des instructions issues de divers langages de programmation peuvent-elles être comprises par le navigateur ?

¹⁹HTML : Hyper Text Markup Language

²⁰ WYSIWYG : What You See Is What You Get

La réponse se trouve dans le mode de fonctionnement d'Internet et dans son architecture client/serveur. Lorsque que l'on entre une adresse dans son navigateur ou que l'on clique sur un lien, notre ordinateur – le client – envoie une requête au serveur sur lequel se trouve la page à laquelle on tente d'accéder. Ce serveur web renvoie la page HTML désirée à notre ordinateur, puis cette page est interprétée par le navigateur et le résultat final s'affiche sur notre écran. C'est ainsi que les choses se passent avec les pages statiques. Dans le cas des pages dynamiques, lorsque que l'on sollicite une page, que notre ordinateur envoie une requête au serveur, le serveur exécute ou traduit le code de notre page dynamique. C'est-à-dire que le serveur a été configuré de manière à être capable d'interpréter le langage dans lequel a été écrite la page dynamique, et à la place de renvoyer cette page de code directement à l'ordinateur client, il traite les instructions de programmation contenues dans cette page et renvoie le résultat final au navigateur sous forme de page HTML.

Les deux schémas ci-dessous illustrent ces 2 situations:



Illustrations extraites de : Online : the leading magazine for information professionals, Volume 24, No. 3, 2000, p. 89

Un exemple pratique de page dynamique illustré par ce schéma pourrait être les dépêches qui figurent sur les premières pages des grands sites d'actualité en temps réel (de type CNN²¹). Le code de la page dynamique va chercher les nouvelles dépêches entrées dans la base de données et le serveur se charge de générer une page HTML à partir de ces dépêches et du code de la page dynamique, qui spécifie la mise en page et la présentation.

Lorsqu'une page est générée par le serveur au moment de la requête de celle-ci par un client, comme c'est le cas avec les pages dynamiques, on qualifie alors le langage utilisé pour réaliser cette page, de langage "server side" (côté serveur en français). Cette terminologie de "côté serveur" est utilisée car la totalité des opérations de génération de la page sont effectuées sur le serveur. Du côté de l'ordinateur client, il ne se passe rien du tout, le client se contente d'envoyer des requêtes au serveur, de recevoir et d'afficher les pages HTML en retour.

C'est, par exemple, le serveur qui ira rechercher des informations dans une base de données, et qui les mettra en forme correctement afin de les envoyer à l'ordinateur client.

Ces opérations relativement élaborées, ont une contre-partie technique assez lourde, il n'est pas possible de créer une page dynamique et de la faire fonctionner sur n'importe quel serveur.

Afin de comprendre à quel niveau agissent les pages créées grâce à ces langages fonctionnant côté serveur, il peut être bon de rappeler succinctement le schéma logique de fonctionnement d'un serveur Internet :

Niveau de fonctionnement	Exemple de produit existant
Système d'exploitation	Windows NT, UNIX, Linux
Logiciel serveur	Apache, IIS
Langage "côté serveur" utilisé pour réaliser les pages	ASP, PHP, Perl

- Le système d'exploitation est le logiciel qui permet à la machine faisant office de serveur de fonctionner. Ce n'est pas spécifique au serveur web, chaque ordinateur a besoin d'un système d'exploitation.
- Le logiciel serveur est le cœur de toute l'opération, c'est le programme qui rend disponible l'information présente sur la machine qui joue le rôle de serveur à d'autres ordinateurs.
C'est la configuration de ce logiciel serveur qui déterminera si telle ou telle page dynamique pourra fonctionner sur le serveur. Par exemple une page créée en utilisant le langage ASP (le langage de Microsoft pour pages dynamiques) ne fonctionnera qu'avec le logiciel serveur de Microsoft (IIS – Internet Information Server).
- Le langage de programmation des pages dynamiques se situe au dernier niveau logique de fonctionnement.

Les pages écrites en HTML sont en revanche compatibles avec tous les types de serveurs, peu importe que le serveur fonctionne avec telle ou telle application, une page HTML sur un serveur web sera toujours lisible, peu importe le fonctionnement de celui-ci.

Ce manque de compatibilité des pages dynamiques, la nécessité de devoir se préoccuper d'éléments aussi techniques que le fonctionnement d'un serveur Internet est due à la complexité des pages dynamiques. Ces dernières ne devraient plus être considérées

²¹ <http://www.cnn.com>

comme de simples documents, mais plutôt comme des programmes à part entière qui sont conçus pour générer des documents. Il est donc nécessaire de s'assurer que ces programmes pourront fonctionner correctement, de la même manière qu'avant d'acheter un CD-Rom, il est nécessaire de vérifier si celui-ci est compatible avec notre type de micro-ordinateur (PC ou Macintosh).

2.3.4 Pourquoi utiliser des pages dynamiques ?

Les raisons de cette évolution, des simples pages statiques au profit des pages dynamiques sont multiples :

- La mise à jour automatisée

La maintenance d'un site composé de pages HTML est très lourde, il faut régulièrement trier et réorganiser chaque page, afin de supprimer les informations périmées. Grâce aux pages dynamiques, il est possible d'instaurer une date de pérennité pour chaque information, le programme comparera ensuite la date de validité d'un document à la date du jour, et supprimera les informations obsolètes. Il est également possible d'entrer des informations à l'avance, informations qui ne seront affichées qu'à partir d'une date définie au moment de l'entrée des données.

- Une séparation logique entre présentation graphique et contenu de la page

En effet, avec des pages HTML statiques (comme c'est le cas avec un document Word), le message, l'information est liée inextricablement à sa présentation graphique. Dans le cas des pages dynamiques, l'information se trouve d'un côté (stockée dans une base de données) et le style, ou la présentation graphique à appliquer à chacune des différentes informations d'un autre. Ce modèle logique permet entre autre une mise à jour simplifiée, dans le sens où l'on ne fait que modifier des données brutes au sein d'une base de données.

Cette séparation des données et de la présentation graphique offre le même confort de gestion que les feuilles de style : il n'est plus nécessaire de modifier un détail de présentation (comme un en-tête par exemple) dans chaque document, mais il suffit de modifier le modèle de document, la présentation type, pour que les changements soient ensuite répercutés sur l'entier du site web.

- Des possibilités d'interaction quasiment illimitées

Les pages dynamiques offrent la possibilité de créer des interfaces permettant n'importe quel type de transaction. On peut par exemple créer une interface de mise à jour des données directement à partir du navigateur. On pourrait tout à fait imaginer un module de catalogage à partir d'un navigateur web pour alimenter une base documentaire. Les forums de discussion que l'on peut voir sur certains sites sont un autre exemple d'utilisation des pages dynamiques. D'une manière générale, il est possible de concevoir n'importe quel type d'application grâce à des pages dynamiques, la technique est la seule limite.

Toutes les perspectives que nous laissent entrevoir les pages dynamiques ont, bien entendu, un prix : l'apprentissage d'un langage de programmation. Même si ces techniques se démocratisent de plus en plus, il est toujours nécessaire d'avoir une certaine connaissance – même minimum - du langage grâce auquel nos pages dynamiques seront créées.

Cependant durant les années à venir on verra arriver de plus en plus de logiciels de création de site web offrant les mêmes possibilités d'interaction que les pages dynamiques, et ceci sans avoir besoin de faire de la programmation. Nous sommes actuellement dans une phase de transition, d'évolution, mais il est certain que de plus en plus de sites seront à l'avenir gérés grâce à des systèmes complexes et que la solution consistant à gérer séparément chacune des pages de son site en HTML sera réservée aux sites de petite taille ou ne nécessitant qu'une mise à jour très peu fréquente.

2.3.5 Les langages de programmation de pages dynamiques : un aperçu

Il ne s'agit nullement ici de faire une présentation exhaustive ou avancée techniquement des divers langages de programmation permettant la création de pages dynamiques, mais plutôt de présenter succinctement quelques langages courants.

La grande majorité des langages de programmation peuvent être utilisés pour interagir (Java, Perl, C Shell, Visual-Basic etc...) avec des pages web, cependant seule une petite partie d'entre eux ont véritablement été conçu à cet effet.

ASP (Active Server Pages), Perl et PHP (Personal Home Page) sont trois langages assez répandus et très couramment utilisés sur le web.

ASP et PHP sont des langages uniquement dédiés au développement de page web dynamiques, contrairement à Perl qui est un langage général (il permet de réaliser n'importe quel type d'application, sans liens avec le web).

Quelques aspects de ces langages "server side":

■ ASP

C'est la solution de Microsoft pour les pages dynamiques. Un nombre de plus en plus important de sites sont maintenant gérés grâce à l'ASP.
Les pages ASP ne fonctionnent qu'en association avec d'autres produits Microsoft.

■ PHP

PHP présente l'intéressante caractéristique d'être gratuit. C'est également le plus simple des trois langages présentés ici. Malgré sa relative simplicité, en comparaison avec un langage comme Perl, il reste un outil très puissant, permettant de réaliser un très grand nombre d'opérations. PHP est notamment un très bon moyen d'interagir avec de nombreux types de bases de données.

■ Perl

Robuste et extrêmement puissant, Perl est un langage fonctionnant sur tous les types de serveurs. Il est utilisé par de nombreux développeurs pour la création d'applications importantes. La complexité de ce langage et son mode de fonctionnement (nettement moins simple que l'ASP ou le PHP) en font un outil uniquement destiné aux professionnels.

2.4 Le web et l'emploi

Les recherches réalisées dans le cadre de mon mandat ont principalement porté sur le domaine de l'emploi. Ce chapitre (le dernier de la partie théorique) est donc consacré au web et à l'emploi.

Quel rôle joue réellement le web dans le marché actuel de l'emploi ? Représente-t-il une révolution ou n'est-il qu'un simple moyen de communication supplémentaire ? En pratique, quel genre de services les sites web concernant l'emploi offrent-ils ? Ces questions seront abordées au cours de ce chapitre.

Ce marché est composé de différents acteurs, et s'il l'on tente de représenter ceux-ci de manière simple et schématique possible, il est possible de limiter ces acteurs au nombre de trois :

- **Le candidat**, à la recherche d'un emploi
- **L'entreprise**, à la recherche d'une personne possédant les compétences nécessaires aux activités de cette entreprise.
- **L'intermédiaire**, favorisant la rencontre du candidat et de l'entreprise.

L'influence d'Internet s'exerce dans la façon dont entreprises et candidats vont entrer en contact. Cette prise de contact peut être directe (le candidat s'adresse à l'entreprise), mais elle passe parfois par un intermédiaire. Ce sont les méthodes de travail de ces intermédiaires qui vont se transformer. Ces intermédiaires sont les agences de placement et les cabinets de recrutement.

Les agences de placement et les cabinets de recrutement sont au cœur de l'activité de médiation dans les rapports entreprises-candidats. Le changement qu'apporte Internet sur cette activité de médiation n'est pas un bouleversement fondamental, il s'agit plutôt d'un moyen supplémentaire pour ces intermédiaires de remplir leur tâche.

2.4.1 Les nouveaux outils de l'emploi

De nouveaux outils ont donc été mis en place, sous la forme de sites Internet. Ces outils permettent la communication et la prise de contact, ils ne correspondent qu'à la première partie d'un processus de sélection et d'engagement divisé en quatre étapes :

1. Un tri selon des critères simples (par exemple sur la base d'un CV).
2. Un second tri basé sur des tests visant à déterminer les compétences.
3. Un contact téléphonique.
4. Un entretien.

Pour l'instant le web permet assez bien de remplir la première étape de sélection, les autres étapes sont plus difficilement réalisables par ce biais.

La publication d'offres d'emploi via le web

Le "nerf de la guerre" de l'emploi sur Internet est bien entendu les offres d'emploi. Internet représente un support idéal à la diffusion de celles-ci. Les offres d'emploi ont été la première forme d'interaction entre Internet et le monde de l'emploi. Les sites leader

dans le domaine de l'emploi ont souvent acquis ce statut grâce à leur rôle de "gisement" d'offres d'emploi.

Dans le domaine des annonces sur le web deux stratégies s'opposent :

- D'un côté on trouve des sites publiant des offres d'emploi issues des sources traditionnelles (la presse ainsi que les cabinets de recrutement).
- De l'autre des sites offrant la possibilité aux entreprises de publier des annonces sur le web à des prix très attractifs.

Pour illustrer cette situation, on citera deux sites français importants :

Le site "Monster"²² étant leader sur le second type de marché, tandis que "Cadre on line"²³ est un acteur très important dans la publication des annonces issues de la presse, avec notamment la reprise des annonces du Monde et du Figaro.

Les entreprises utilisent parfois leur propre site web pour publier leur offre d'emploi, toutefois l'efficacité de ce genre de démarche n'est pas forcément optimale : en effet, les offres ainsi publiées ne bénéficient pas d'une diffusion très large. Mais cette tendance est à relever.

Le dépôt de CV et la candidature via le web

Les CV et présentations de candidats grâce au web sont représentatives des espoirs qu'a suscité Internet dans le monde de l'emploi : Internet a été vu comme la panacée, comme une solution qui permettrait de réunir entreprises et candidats de manière simple, en supprimant presque tout intermédiaire.

Les entreprises de leur côté, se sont mises à publier leurs offres d'emploi sur différents sites, ceci à des prix très bas (sans comparaison avec les moyens traditionnels tel que la presse).

Tandis que les candidats inscrivent gratuitement leur CV dans des "banques de CV". Les sites offrant cette possibilité sont des grands réservoirs de candidat, que les entreprises viennent consulter quand elles sont à la recherche de personnel.

Cette situation qui se voulait idéale débouche en fait sur une impasse :

Le nombre de candidats proposant leurs CV est beaucoup trop élevé, et leurs profils sont très hétérogènes, le processus de recherche d'un candidat par une entreprise se complique alors considérablement... et le miracle attendu d'Internet n'a pas lieu.

C'est pour cette raison que les sites répertoriant simplement des CV, sans offrir de services supplémentaires ont tendance à se faire plus rares.

La tendance actuelle est plutôt d'augmenter la qualité des candidats (via une sélection stricte) afin de diminuer le nombre de candidatures et de présenter aux entreprises des candidats ayant des profils correspondant à leurs besoins (bien entendu ce service à un prix).

Pour illustrer l'échec des services de type "dépôt de CV", on peut citer l'exemple du site de l'ANPE Française :

²² <http://www.monster.fr/>

²³ <http://www.cadreonline.com/>

Ouvert en 1997, le but principal du site de l'ANPE est de diffuser la liste des emplois disponibles de l'ANPE. Ce site offrait également un service de dépôt de CV, mais suite au constat de l'inefficacité de ce genre de démarche ce service a été modifié. Il n'a pas été totalement supprimé, mais le dépôt de CV est maintenant réservé à des cas bien spécifiques (jeunes diplômés ou candidats désirant travailler à l'étranger).

2.4.2 Le web et l'emploi en Suisse romande

Malgré une assez forte utilisation du web en Suisse (44% des suisses déclarent avoir déjà surfé) les synergies entre Internet et l'emploi ne sont pas au plus haut.

A l'inverse de la France, on ne constate pas en Suisse l'arrivée massive de nouveaux acteurs dans le domaine de l'emploi sur Internet. Il existe en Suisse encore très peu de sites assumant un rôle d'intermédiaire entre candidats et entreprises. Il ne s'agit pas pour autant de déclarer qu'il ne se fait rien en matière d'emploi dans le web suisse ! Les sociétés actives sur le web sont simplement des acteurs "traditionnels" qui prolongent leur activité habituelle dans le cadre du web. Les agences de placement et d'intérim offrent par exemple très souvent des offres d'emploi via leur site, le groupe Edipress republie l'ensemble des offres d'emploi parues dans ses quotidiens et autres titres de presse.

3 Réalisation du projet

3.1 Genèse et objectif du projet

J'ai rencontré Monsieur Daniel Porot, le mandataire de mon travail de diplôme, dans le contexte du projet " Boutique Emploi " :

L'idée de base de la Boutique Emploi est la suivante : créer un lieu de rencontre entre chercheurs d'emploi et entreprises, afin de faciliter l'insertion des personnes à la recherche d'un emploi – les candidats - dans le tissu économique. Pour amener ces deux parties à se rencontrer dans ce lieu, il faut parvenir à les y intéresser. Pour ce faire, la Boutique Emploi offre divers services aux candidats et aux entreprises.

L'un des ces services étant notamment un site web. Ce site web doit refléter cette approche de service aux deux publics cibles, et d'être un premier lieu de rencontre (même virtuel) entre candidats et entreprises.

Même si les deux publics se rendent sur le site pour y chercher chacun une information différente, ils fréquentent le même lieu. Le contenu du site est en quelque sorte un produit d'appel, un moyen d'approche possible de la Boutique Emploi pour les deux publics cible.

Le service offert par ce site web est donc le suivant : donner un accès simplifié aux ressources (susceptibles d'être utiles aux candidats et entreprises) disponibles sur le web.

Il s'agit d'exploiter la richesse de l'information présente sur Internet, richesse brute et dispersée, pour créer un produit, un répertoire de ressources attractif.

Dans le contexte de ce projet, mon mandat avait donc pour but la sélection des différents sites et pages Internet formant le contenu du site de la Boutique Emploi, et la réalisation technique de ce site.

Ce projet - la Boutique Emploi – est en cours de création, mais sa mise sur pied prend plus de temps que prévu.

Suite au retard dans la réalisation du projet de la Boutique Emploi, mon mandat a été intégré à une autre activité du cabinet Daniel Porot à savoir le Club Emploi.

Le Club Emploi est une structure offrant aux personnes à la recherche d'un emploi, un lieu de travail et d'assistance pour leur permettre de retrouver rapidement un emploi leur correspondant.

Le Club Emploi fonctionne de la manière suivante : chaque personne inscrite accepte de consacrer au Club Emploi, en moyenne deux demi journées par semaine et ce pendant 12 semaines au maximum (ou moins en cas de reprise d'emploi).

Le Club Emploi offre aux personnes inscrites les moyens matériels (téléphones, ordinateurs, bureaux) de réaliser leur recherche d'emploi, tous les vecteurs possibles d'offres d'emploi (tel que petites annonces de quotidiens etc...), des ateliers de formation, des cours, de la documentation sur les sujets suivants : les techniques d'entretien, la définition des objectifs, la création de réseaux, ainsi qu'une aide personnalisée à la réalisation de documents tels que : dossier, réponse de petite annonce, CV, lettre de motivation.

Il existe à l'heure actuelle un Club Emploi dans les villes suivantes : Genève, Lausanne, Sion, Fribourg.

L'objectif du mandat concernant le Club Emploi est donc le même que celui qui aurait dû être réalisé dans le cadre de la Boutique Emploi, à savoir : exploiter les possibilités offertes par Internet dans les domaines concernant le Club Emploi afin de créer un outil répertoriant des ressources Internet. Cet outil prenant la forme d'un annuaire de type Yahoo! – mais focalisé sur le domaine de l'emploi en suisse romande.

En pratique, ce répertoire de site sera utilisé par les participants du Club Emploi. Cependant le projet de la Boutique Emploi étant toujours à l'ordre du jour, le choix a été fait de conserver une double approche candidat-entreprise. C'est-à-dire, de diviser le répertoire en deux et de présenter aussi bien des informations pour les candidats, que pour les entreprises. Ceci pour conserver la possibilité de faire évoluer l'outil vers ce qu'il aurait dû être à la base, c'est-à-dire le site web du projet "Boutique Emploi".

3.2 Le mandat

Mon mandat définitif est donc le suivant :

- **L'élaboration d'un répertoire proposant des ressources présentes sur Internet dans le domaine de l'emploi et de la formation. Ce répertoire utilisera une double approche : il sera destiné aux candidats comme aux entreprises.**

Le répertoire a deux domaines d'application :

- Le Club Emploi

La partie du répertoire destinée aux candidats sera utilisée par les participants au Club.

- La Boutique Emploi

Une fois celle-ci mise en place, le répertoire constituera une base pour le site de la Boutique Emploi

A ces objectifs de base ont été ajoutées les contraintes suivantes :

- L'outil final sera d'usage simple.
- La mise à disposition du répertoire pour les participants au Club Emploi sera réalisée par le biais d'un site protégé par mot de passe.
- La mise à jour et la gestion de l'outil devront être réalisable de manière simple et aisée.
- Le champ couvert par le répertoire est centré sur les ressources Internet en Suisse romande.
- La sélection des ressources sera très fortement focalisée sur l'emploi, c'est-à-dire des informations sur des entreprises, des sites aidant à la recherche d'emploi et à la carrière, les gisements d'annonces (offres d'emploi) et enfin des sites recensant des candidats (banques de CV).
- Une partie du travail de recherche de sites a déjà été effectuée dans le cadre d'un premier mandat de 40 heures par M. Nicolas Derron. Le début de la collecte d'information sera basé sur l'arborescence et le travail déjà effectué.

■

Quel outil ?, Pour quel public ?

La fonction de service remplie par le répertoire est la suivante :

- Rendre l'information disponible sur Internet concernant l'emploi en Suisse romande aisément accessible aux participants du Club Emploi.

Une fois cette base établie, il est nécessaire de déterminer le public du Club Emploi. Ceci pour adapter l'outil au public cible.

Ce public est assez hétérogène, il peut s'agir d'employés, de techniciens comme de cadres. Toutes ces personnes effectuent une démarche de recherche d'emploi, elles peuvent avoir une connaissance ou une habitude d'utilisation très variable d'Internet.

Le but étant de fournir un service le plus "démocratique", le plus accessible possible, la démarche ne sera pas de déterminer une connaissance moyenne au sein du public cible et de créer un outil correspondant à cette connaissance moyenne, mais plutôt de viser la plus grande simplicité possible dans la conception de l'outil, ceci pour permettre au plus grand nombre de personnes possible de l'utiliser sans difficulté.

Il ne faut cependant pas se fixer de but utopique. Même si la simplicité est un des objectifs de base, il faut être conscient que quel que soit l'ergonomie et l'accessibilité de l'outil, il représentera toujours une difficulté pour une personne n'ayant absolument aucune habitude ou expérience d'utilisation d'un ordinateur.

Le niveau de simplicité optimal du répertoire pourrait être déterminé ainsi: si une personne ayant une petite habitude d'utilisation d'un ordinateur, *mais n'ayant jamais surfé sur Internet*, est capable d'utiliser le répertoire, on peut considérer l'objectif de simplicité comme atteint.

3.3 Méthodologie de travail

Chronologiquement mon travail s'est divisé en trois phases distinctes :

- La première était dédiée à la collecte d'information. Durant cette première étape je ne me suis pas préoccupé de la forme de l'outil final ou de considérations techniques, mais uniquement de réunir et de recenser la matière brute du site.
- Durant la seconde phase, j'ai pratiquement cessé la recherche d'information pour me concentrer sur la forme de l'outil final. Il a fallu déterminer et mettre en place la solution informatique permettant la diffusion et la gestion des informations recensées.
- La dernière phase ou finalisation m'a permis de combiner les résultats de recherche avec la solution informatique et de présenter le produit fini.

Les deux enjeux importants de ce travail étaient la recherche d'information sur le web et la diffusion des résultats de cette recherche par le biais d'un système en ligne.

Le mandat imposait des contraintes à deux niveaux différents :

- Une contrainte de contenu au niveau de la recherche (l'emploi et la formation en Suisse romande)
- Une contrainte de forme au niveau de la solution informatique (la simplicité).

3.3.1 Méthodes de recherche utilisées

Mon champ de recherche était défini par le mandat; il s'agissait des sites concernant l'emploi et la formation. Ces sites pouvant s'adresser aux candidats comme aux entreprises.

Ce champ de recherche avait déjà été cerné par M. Nicolas Derron. Son travail a donc constitué une base de départ pour mes propres recherches.

La particularité de ce projet par rapport à un travail documentaire "classique" est importante : à la place d'être considéré comme une source d'information parmi d'autre, **Internet était la seule source d'information.**

Cette particularité est importante car elle représente une prise de risque. En effet, Internet n'offre pas les garanties de qualité et de structuration du monde documentaire "traditionnel". En plus de cette absence de garantie concernant la qualité de l'information, une seconde incertitude était présente au niveau de la quantité d'information disponible : le champ de la recherche ne recouvre pas l'entier d'Internet, mais principalement le web en Suisse romande.

Mon approche de recherche a été la suivante :

Elaborer des listes de mots-clés à partir des thèmes et sujets se trouvant dans mon champ de recherche, puis sur la base de ces listes de mots-clés effectuer des requêtes sur différents moteurs de recherche.

La plus grande partie de sites web suisses utilisent un nom de domaine terminé par ".ch" (une autre partie, moins importante, utilise les nom de domaines en ".com" ou ".net"). Pour atteindre le web suisse, il m'a fallu sélectionner les outils capables de filtrer les

résultats par nom de domaine (.ch). Ceci a assez fortement limité la gamme d'outil à ma disposition :

- Tous les moteurs de recherche incapables de filtrer les résultats n'ont pas été utilisés.
- Les annuaires "non-suisse" n'ont pas été utilisés car ils ne contiennent évidemment pas d'information concernant la Suisse.

J'ai donc utilisé les moteurs de recherche suivants :

Nom du moteur	Adresse URL
Altavista	http://www.av.com
Fast	http://www.alltheweb.com
Google	http://www.google.com
Hotbot	http://www.hotbot.com
Infoseek	http://www.infoseek.com
Northernlight	http://www.northernlight.com
Sear.ch (Altavista suisse)	http://www.sear.ch

Il est nécessaire de distinguer les deux étapes d'une démarche de recherche:

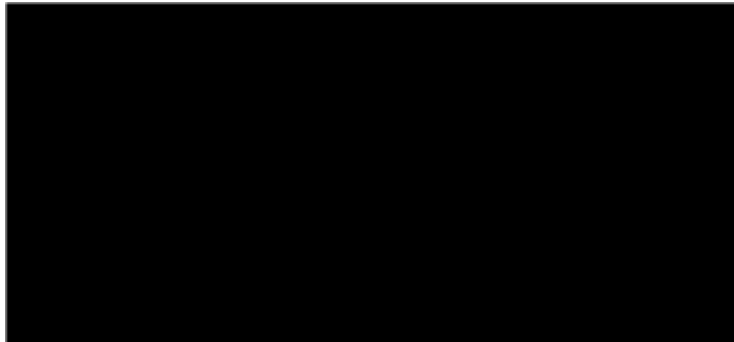
- La recherche active, par l'interrogation des outils sélectionnés
- Le "surf" ou l'exploration des liens hypertextes (de page en page)

Ces deux manières de procéder ne s'excluent pas, elles sont bien les deux étapes d'une seule et même démarche. La méthode consistant à explorer les liens à partir d'un site donné est très productive et permettra la découverte d'une grande quantité d'information pertinente. Grâce à cette démarche de "surf", on obtiendra assez rapidement des résultats, tandis qu'une recherche effectuée par l'intermédiaire d'un moteur peut s'avérer longue et parfois peu fructueuse.

Cette situation est normale, le surf constituant la deuxième étape d'une démarche de recherche. Le rôle attribué aux outils de recherche étant de nous donner un point de départ dans l'exploration des liens. Bien entendu plus ce point de départ est pertinent est plus la recherche sera rapide et fructueuse.

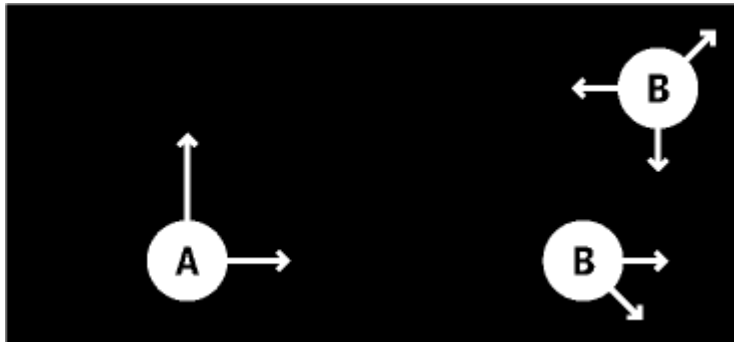
Les schémas suivants illustrent cette situation :

1



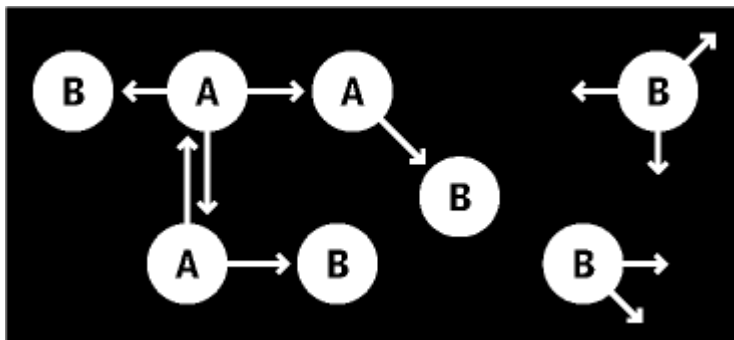
Au départ de la recherche, on ne dispose d'aucune information, aucun repère n'est disponible.

2



Après interrogation d'un outil de recherche, un certain nombre de sites sont découvertes ("A" représente un site pertinent et B un site non pertinent). Ces sites pointent eux-même sur d'autres sites.

3



Ce dernier schéma illustre la quantité d'informations pertinentes découverte après une phase de "surf" ou d'exploration systématique des liens. La quantité d'informations pertinentes mises à jour dans cette phase est plus importante que dans la première phase de recherche réalisée en interrogeant un outil.

Dans une recherche, il faut également tenir compte du fait que toutes les informations et tous les sites présents sur le web ne sont pas sur un pied d'égalité : certaines

informations sont plus aisées à trouver que d'autres.

Le référencement est un facteur très important : un site bien référencé dans les outils de recherche sera facilement accessible. Un référencement efficace doit être effectué par des spécialistes de ce domaine, et ce travail coûte cher. Ceci a pour conséquence de rendre plus disponibles les sites des organismes ayant suffisamment de moyen pour s'offrir un référencement efficace.

3.3.2 Résultats des recherches

Je ne vais pas faire ici de liste de sites ou de pages. Pour consulter les résultats de la recherche, on consultera le répertoire à l'adresse suivante :

<http://www.clubemploi.ch>

Nom d'utilisateur : internet
Mot de passe : internet

Concernant la quantité de sites obtenus, on peut constater un déséquilibre entre les deux différentes sections du répertoire :

La section entreprise est nettement moins développée que la partie candidat, et pour cause : elle ne sera pas utilisée par les participants du Club Emploi. Le but étant pour eux de trouver un emploi, j'ai donc principalement concentré mes recherches sur la section candidat.

Les constatations suivantes sont à relever concernant les différents domaines du champ de recherche :

- Formation

Le web est un domaine de prédilection pour la formation, un très grand nombre d'écoles et d'institutions assurant des formations sont présentes sur le web. Les universités et les écoles polytechniques fédérales sont notamment des exemples d'exploitation du web en tant que moyen de communication et d'information. De plus la Confédération, qui est en Suisse, une des meilleures sources d'information sur le web, joue un rôle centralisateur en ce qui concerne la formation (par exemple pour tout ce qui touche à l'apprentissage).

- Emploi

L'emploi est un domaine relevant tant de l'économie et du domaine commercial (les entreprises) que du service public (l'Etat jouant un rôle d'insertion des candidats dans le tissu économique, notamment par le biais des ORP – office régional de placement). Cette dualité se retrouve dans les différents sites ayant trait à l'emploi:

Les sites commerciaux (par exemple des offres d'emploi ou des répertoires d'entreprises) ne contiennent généralement que peu de liens vers d'autres sites dans le même domaine. Ceci s'explique aisément : une entreprise n'a aucun intérêt à établir des liens vers d'autres sites, ceux-ci représentant généralement une concurrence.

Par contre les sites de la Confédération et des Cantons concernant l'emploi et le chômage tentent de donner accès à un maximum de ressources et d'offrir le plus de liens possible. Cette volonté de créer des liens a pour effet de donner (dans ce domaine) une simili-structure au web. La recherche est simplifiée et les informations

sont disponibles plus facilement. Tandis que les informations sur l'emploi relevant du secteur privé sont plus difficiles à atteindre.

Certains sites d'associations ou de fédérations jouent un rôle de site carrefour : ils recensent et font des liens vers un grand nombre d'autres sites. On citera à titre d'exemple le site de la Fédération des Syndicats Patronaux²⁴.

3.3.3 Évaluation et sélection des sites

Il existe de nombreux critères et grilles de sélection pour évaluer l'information présente sur internet (cf bibliographie). Certains éléments importants se retrouvent pratiquement systématiquement dans les différentes méthodes d'évaluation de l'information disponible sur Internet.

J'ai construit deux grilles d'évaluation à partir de ces éléments.

Voici les grilles de sélection et d'évaluation utilisées :

Critères de sélection	Évaluation
Pertinence	Le sujet de l'information, son contenu entre-il dans le champ de la recherche ?
Langue	L'information est-elle en français ?
Couverture géographique	L'information concerne-t-elle la Suisse romande ?

Critères d'évaluation	Évaluation
Origine	Qui est l'auteur d'une information, est-ce une personne privée ou une institution ? Cette institution est-elle reconnue ?
Mise à jour, fraîcheur de l'information	Quelle est la date de la dernière modification ? La mise à jour est-elle un facteur important pour ce type d'information ?

Remarques concernant la procédure de sélection :

- Les informations dont il n'était pas possible de déterminer l'auteur ou la source n'ont généralement pas été retenues.
- 98 % des sites sélectionnés sont en français, toutefois quelques sites en allemands ou en anglais particulièrement intéressants figurent dans le répertoire. Dans ce cas, il est fait mention de langue dans la description du site.
- Toutes les informations n'ont pas nécessairement besoin d'être mises à jour à des fréquences élevées. La mise à jour n'est donc pas forcément un élément discriminant.

²⁴ <http://www.fsp.ch>

3.3.4 Structuration des résultats : L'arborescence du répertoire

La solution choisie pour structurer les résultats de la recherche est une arborescence à trois niveaux.

Les deux sections du répertoire (respectivement Candidat et Entreprise) sont divisées en plusieurs catégories de base. Ces catégories sont les premières informations affichées lorsque l'on commence à consulter le site. Chaque catégorie de bases peut ensuite contenir jusqu'à deux "sous-catégories".

L'arborescence complète figure dans les annexes.

L'arborescence simple a été choisie comme solution de classement car elle correspond aux deux contraintes définies par le mandat :

- Simplicité d'utilisation

L'arborescence offre une approche simple et intuitive de l'outil, le choix d'une arborescence peu complexe, limitée à trois niveaux, en renforce encore la simplicité. Le nombre réduit de niveaux d'arborescence permet d'appréhender rapidement la totalité du contenu du répertoire.

En revanche, la quantité d'information gérable grâce ce système est limitée et au-delà d'un certain nombre de sites l'objectif de clarté n'est plus atteint, car il y a trop d'information par catégorie. Toutefois cette solution est tout à fait adaptée à la taille actuelle et future du répertoire. (Le web suisse romand n'étant pas d'une taille énorme).

- Simplicité de gestion, de mise à jour.

Le classement par arborescence permet également une gestion simple : la seule opération "d'indexation" nécessaire pour chaque site est de lui attribuer une catégorie. Chaque site ne peut donc être présent que dans une seule et unique catégorie. Bien que cette solution implique une gestion simple, il aurait été préférable de pouvoir attribuer plusieurs catégories à chaque site. Ceci aurait permis de nuancer l'indexation, et de retrouver les sites plus aisément. Malheureusement la solution technique choisie n'offrait pas cette possibilité.

Le choix des catégories a été réalisé à la fin du processus de recherche. Les catégories ont donc été choisies pour structurer le mieux possible les résultats de la recherche. Le contenu a déterminé l'arborescence.

Cette arborescence, tout comme un thésaurus ou n'importe quel autre système de description de l'information, est appelée à évoluer parallèlement au contenu. Ce besoin d'évolution a été pris en compte et la solution technique utilisée pour gérer le répertoire permet de créer de nouvelles catégories et de déplacer les sites d'une catégorie à l'autre.

3.4 Réalisation technique

La partie technique de ce travail de diplôme - le choix d'une solution et la mise en place de celle-ci - était importante.

Le mandat définissait les deux conditions suivantes pour la solution technique :

- Etre simple d'utilisation
- Permettre une mise à jour aisée, ne requérant pas de compétence particulière.

Il n'y avait pas de cadre informatique préexistant pour le Club Emploi ou pour la Boutique Emploi, ma liberté de choix concernant la solution était donc très large.

Il aurait été possible de fournir les résultats sous forme de pages HTML. Mais cette solution ne remplissait pas réellement la deuxième condition fixée par le mandat. Et si le répertoire avait été réalisé en pages HTML, le travail n'aurait pas été très différent d'une simple bibliographie sur l'emploi. Sa seule particularité aurait été de recenser uniquement des sources Internet. Les potentialités d'interactivité offertes par Internet seraient alors restées inexploitées.

La seconde condition précisée par le mandat est importante car une mise à jour simple implique la mise en place d'une véritable application et non pas l'utilisation de pages HTML.

Le défi était donc de fournir un système de gestion du répertoire répondant aux contraintes de simplicité et de mise à jour. Pour remplir ces contraintes le système devait fonctionner au moyen de pages dynamiques.

Pour mettre en place un tel système, il était possible d'acheter une solution commerciale ou d'adapter une solution disponible gratuitement sur le web. J'ai opté pour l'adaptation d'une solution gratuite, pour la flexibilité offerte par celle-ci. En effet les solutions gratuites sont généralement « open source » ce qui signifie qu'il est possible d'y apporter toutes les modifications souhaitées (dans les limites de ses propres compétences informatiques).

Choisir d'adapter une solution préexistante nécessite un minimum de maîtrise du langage ayant été utilisé pour créer cette solution. Avant de commencer ce travail de diplôme je n'avais jamais travaillé avec des pages dynamiques ou réalisé d'application fonctionnant sur le web.

Avant de rechercher un programme capable d'assurer la gestion du répertoire, il était nécessaire de choisir parmi les différentes techniques permettant de réaliser des pages dynamiques (ASP, Perl, PHP etc...).

Dans ce contexte, le choix du langage PHP est apparu comme une excellente solution pour les raisons suivantes :

- De nombreuses solutions pour gérer des annuaires ou répertoires de liens utilisant le PHP sont disponibles gratuitement.
- PHP est un langage rapide et puissant mais qui reste relativement simple; son apprentissage en autodidacte durant mon travail de diplôme était donc possible.
- Une application réalisée à l'aide de PHP fonctionne sur la plupart des serveurs web.

- Etant donné la gratuité de ce langage, et le fait qu'il ne soit pas lié à une société qui le commercialise, de nombreuses ressources sont disponibles sur le web (aide en tout genre, forums de discussion, exemples d'applications).
- PHP est un langage conçu spécifiquement pour réaliser des applications fonctionnant sur le web. Il est très souvent associé à un système de base de données (également destiné au web, et également gratuit : MySQL).

Choisir d'adapter une solution gratuite peut apparaître comme une solution peu professionnelle, de plus la gratuité du langage PHP peu également lui conférer une image d'outil peu fiable ou destiné à des passionnés du genre désireux d'expérimenter : il n'en est rien. PHP est un langage utilisé par de très nombreuses entreprises et organismes pour gérer leur site web (et notamment la compagnie Ericsson²⁵, ou encore le journal El Mundo²⁶).

Lors de mon stage de formation de 2^{ème} année au Service Documentation et Archives (D+A) de la Télévision Suisse Romande, une application semblable (un répertoire de site web appelé : "Sitoscope") avait été réalisée par M. Pierre Crevoisier (journaliste à la TSR), cette application avait entièrement été écrite en PHP et utilisait une base de données MySQL.

Savoir qu'une application similaire a été réalisée (et soit parfaitement fonctionnelle) grâce à PHP m'a conforté dans mon choix.

J'ai donc appris les bases du langage PHP, ainsi que les rudiments du système de base de données MySQL pour me permettre de comprendre le fonctionnement des diverses solutions et de pouvoir les évaluer et afin de choisir une solution et d'en assurer la mise en place.

Je n'étais absolument pas certain de réussir la mise en place de l'application, mais si les problèmes techniques s'étaient avérés trop importants, il m'aurait toujours été possible de présenter les résultats de la recherche sous forme de page HTML et de confier la réalisation technique à une tierce personne. Le risque d'un échec au niveau informatique n'aurait pas remis en cause l'intégralité de mon travail de diplôme.

²⁵ <http://www.ericsson.de/>

²⁶ <http://www.el-mundo.es/>

3.4.1 La solution technique choisie : phpHoo 2

J'ai recherché et testé diverses applications écrites en PHP créées pour gérer des répertoires de sites web. Voici une liste des applications que j'ai été amené à tester :

Nom de l'application	Adresse URL
Link conservatory	non disponible
Link base	non disponible
PhpHoo	http://www.webreference.com/perl/xhoo/hp1/
PhpHoo 2	http://www.cable-modems.org/phpHoo/
Phod	http://imexis.com/phod.php3.txt
Php portal	http://mazen.xodox.de

Mon choix s'est finalement porté sur l'application phpHoo 2, car elle offre les fonctionnalités suivantes :

- La possibilité de gérer une collection de sites web.
- La gestion de ces sites web par le biais de différentes catégories et sous-catégories
- La possibilité de commenter chaque site.
- Un module permettant la mise à jour à partir d'une interface web (ajouter de nouveaux sites, éditer les sites présents dans le répertoire, ajouter de nouvelles catégories).
- Un moteur de recherche permettant d'effectuer des recherches par mots-clés.
- Une présentation simple et claire.

De plus, la documentation et les commentaires disponibles avec phpHoo 2 sont tout à fait complets. Ceci représente un avantage important par rapport à d'autres applications qui sont fournies "tel quel" et sans aucun commentaire.

Toutefois, comme toute solution, PhpHoo 2 a ses limitations :

- La première limitation se situe au niveau de la présentation graphique. La mise en page et la manière dont sont présentées les informations sont des éléments définis par PhpHoo 2; il n'est pas possible d'utiliser une autre présentation et de conserver le système "mécanique" qui gère les différentes catégories. La seule façon de modifier la mise en page est de réécrire une partie du programme, c'est un travail long et qui nécessite une bonne connaissance du langage PHP (connaissance que je n'ai pas).
- La seconde limitation est une question d'indexation. PhpHoo ne permet de classer les sites gérés dans le répertoire que dans une seule et unique catégorie. Ainsi pour un site offrant des conseils concernant la rédaction de CV ainsi que des offres d'emploi, il sera nécessaire de choisir entre la catégorie "CV" et la catégorie "Offres d'emploi". Ce site ne pourra pas figurer dans les deux catégories, bien que cela eut été parfaitement justifié d'un point de vue de gestion documentaire.

PhpHoo 2 a été créé par Rolf V. Ostergaard qui en autorise l'utilisation et la modification libre.

Voici un exemple d'utilisation de phpHoo 2 sur le site de son auteur.



Comment fonctionne PhpHoo 2 ?

Les adresses et descriptions des sites présents dans le répertoire ne sont pas stockées dans les pages du site, mais dans une base de données. Lorsque le répertoire est consulté, PhpHoo 2 rapatrie les informations présentes dans la base de données (les noms et descriptions des sites), puis à partir de ces informations, il génère la page HTML qui sera affichée à l'écran.

PhpHoo 2 n'est donc pas un simple programme, mais une combinaison entre un programme et une base de données.

La base de données utilisée en collaboration avec PhpHoo 2 est une base de type "MySQL"²⁷. Les bases de données MySQL sont spécifiquement conçues pour fonctionner sur un serveur web.

A l'inverse des bases de données Access ou FileMaker, les modifications et mises à jour d'une base MySQL s'effectuent directement sur le serveur hébergeant celle-ci et par le biais d'une interface web.

Il n'est donc pas possible de créer une base de données sur un ordinateur non relié à Internet, de la modifier puis de la placer sur un serveur Internet. La base de données doit être créée sur un serveur, puis tous les changements et mises à jour sont à effectuer directement sur le serveur.

L'interface utilisée pour gérer la base de données est l'application "phpMyAdmin"²⁸. Cette application est utilisée via un navigateur web.

²⁷ On trouvera plus d'information concernant MySQL à l'adresse suivante : <http://www.mysql.net>

²⁸ On trouvera plus d'information concernant phpMyAdmin à l'adresse suivante : <http://www.phpwizard.net/phpMyAdmin/>

Accueil
clubemploi

Base de données clubemploi

table	Action	Enregistrements
Categories	Afficher Sélectionner Insérer Propriétés Effacer Vider	31
Categoriесе	Afficher Sélectionner Insérer Propriétés Effacer Vider	9
Links	Afficher Sélectionner Insérer Propriétés Effacer Vider	182
Linkse	Afficher Sélectionner Insérer Propriétés Effacer Vider	60

- Exécuter une ou des requêtes sur la base clubemploi [\[Documentation\]](#):

ou Emplacement du fichier texte:
- [Requête par un exemple](#)
- Afficher le schéma de la base
☒ Structure seule ☐ Ajouter des énoncés 'drop table'
☐ Structure et données ☐ transmettre
- Créer une nouvelle table sur la base clubemploi:
Nom:
Champs:
- [Détruire la base clubemploi](#) [\[Documentation\]](#)

Grâce à cette application, il est possible d'effectuer toute la gestion de la base de données, il est également possible d'effectuer des extractions de données. Le contenu entier du répertoire peut ainsi être sauvé dans un fichier texte. Cette possibilité de sauvegarde est très intéressante, car elle offre également la possibilité de récupérer les données afin de les réutiliser dans un autre contexte.

3.4.2 La mise en place de l'application

Après une première période de test, j'ai finalement installé PhpHoo sur un serveur commercial, ce serveur héberge également le site du Club Emploi.

J'ai toutefois apporté les changements suivants à l'application :

- Traduction des textes en français (tous les textes de l'interface de PhpHoo 2 étant à l'origine en anglais)..
- Suppression de la recherche par mots-clés.

En effet, l'application offrait un champ permettant d'effectuer des recherches, mais il n'a pas été conservé afin de rendre l'utilisation du répertoire encore plus simple. La possibilité de rechercher via des mots-clés ayant tendance à troubler les utilisateurs peu habitués au web. Ils tentent parfois d'utiliser la recherche et n'obtiennent pas de résultat, ce qui les pousse à abandonner.

- Protection du module d'administration par un système de mot de passe.

Le résultat auquel je suis arrivé est un compromis entre une formule qui se doit de rester la plus simple d'utilisation possible et entre un outil performant qui se veut gérable à long terme.

3.5 Utilisation et évolution future du répertoire

Actuellement le répertoire est utilisé par les participants du Club Emploi (à Lausanne, Genève, Sion et Fribourg). C'est pour cette raison que le répertoire a été placé sur le site du Club Emploi.

Deux cas de figure sont maintenant envisageables pour l'évolution future du répertoire :

- Etre intégré tel quel au site de la Boutique Emploi, quand celui-ci sera réalisé.
- L'abandon de la solution technique (PhpHoo 2) au profit d'une solution plus performante (nécessitant des investissements plus importants). Dans ce cas, le contenu du répertoire pourrait être extrait dans son intégralité et réutilisé dans le nouvel outil à venir.

Etant donné la nature versatile de l'information présente sur Internet et recensée par le répertoire, une mise à jour régulière est nécessaire, sous peine de voir le contenu être assez rapidement obsolète.

J'estime le nombre d'heures minimum nécessaire pour tenir le répertoire à jour à environ 10 heures par mois.

Afin de faciliter la gestion du répertoire par une tierce personne, j'ai rédigé deux documents :

- Un manuel d'administration, présentant les possibilités de mise à jour.
- Une notice technique, permettant de comprendre le fonctionnement "mécanique" du répertoire.

Ces deux documents figurent dans les annexes.

4 Conclusion

Tout au long de ce travail, j'ai toujours été partagé entre la réalisation technique, et le travail plus documentaire de recherche et description des sites. C'était pour moi une situation très plaisante d'avoir la possibilité de remplir les deux rôles, de réaliser l'outil dans sa globalité aussi bien au niveau documentaire, qu'au niveau informatique. Bien que cette double réalisation fut très intéressante, elle présente le désavantage suivant : tout ce que j'ai réalisé au niveau informatique, représente autant de chose que je n'ai pas faites au niveau documentaire et vice versa.

Si l'on imagine une situation " normale " de création et développement d'un outil tel que ce répertoire de liens, on aurait certainement eu à faire à plusieurs acteurs, certains remplissant les tâches documentaires et d'autres les tâches informatiques.

Bien que certaines de tâches informatiques réalisées au cours de ce travail de diplôme ne relèvent pas du domaine documentaire, je les estime importantes. Je m'intéresse à l'informatique, c'est un fait, et même si j'ai certaines compétences dans ce domaine, je ne suis pas un informaticien et je ne prétends pas l'être. Dans une situation normale, je n'aurais donc pratiquement rien réalisé de la partie technique. Toutefois les compétences et l'expérience que j'ai pu acquérir dans ce domaine au cours de ce travail de diplôme sont loin d'être inutiles. En effet, avoir des compétences informatiques avancées est important, non pas pour réaliser soi-même des travaux que d'autres plus compétents feront mieux et plus vite, mais pour comprendre les implications techniques inhérentes à un produit, un service ou une application.

Les produits documentaires actuels utilisant de plus en plus les nouvelles technologies, il devient donc capital de pouvoir collaborer avec les personnes concevant la partie technique de ces produits. Et ceci passe par une connaissance – *à jour* - des solutions techniques actuelles. Le domaine étant en constante évolution, il est important de rester à niveau, sous peine de voir un certain nombre de tâches nous échapper.

Bien que ce travail de diplôme n'ait pas été réalisé dans le cadre d'un service de documentation ou d'une bibliothèque, il est bien plus "bibliothéconomique" qu'il n'y paraît de prime abord. En effet, avec le développement d'Internet, et l'accroissement du nombre de services que les SID vont offrir au travers de leur site web, des outils semblables à celui que j'ai été amené à réaliser au cours de ce travail de diplôme vont apparaître de plus en plus fréquemment sur les sites Internet ou intranet des SID. A l'heure actuelle un nombre de plus en plus important de bibliothèques offrent déjà une sélection de ressources Internet en rapport avec les disciplines de leurs fonds (par exemple la Bibliothèque cantonale et universitaire du canton de Vaud²⁹).

Pour l'instant ces sélections de sites sont traitées comme de simples listes et figurent souvent dans les pages "liens" ou dans une section du site intitulée "pour aller plus loin". Mais ce n'est qu'une question de temps avant que ces listes de liens ne deviennent trop importantes pour être gérées uniquement au moyen de pages HTML. La gestion et le développement de ce type d'outil sont donc des compétences qui entrent tout à fait dans le champ des nouvelles fonctions des professionnels de l'information.

Et au-delà du monde strictement documentaire, les répertoires ou annuaires de sites web sont des outils performants qui permettent de répondre à un besoin de plus en plus grand : l'utilisation efficace du potentiel informatif d'Internet. Ces types d'outils et de services vont également être de plus en plus présent dans les portails d'entreprises.

²⁹ <http://www.unil.ch/BCU/recherch/liens.htm>

Ce projet à été pour moi particulièrement formateur, à différents point de vue:

- C'était notamment ma première expérience de réalisation d'un mandat, de prise en charge d'un projet dans son intégralité. Il m'a donc fallu gérer des paramètres tel que le respect des délais ou encore le dialogue professionnel avec mon commanditaire.
- La grande liberté dont je disposais et la réalisation du mandat en dehors du cadre d'une institution étaient également des éléments nouveaux pour moi. Bien entendu, j'étais encadré et assisté de manière excellente par ma directrice de travail Anne-Christine Gourdal mais c'était la première fois que je jouissais d'autant de liberté dans le cadre de mon travail professionnel. Mes conditions de travail étaient réellement celles d'un indépendant :
Je n'ai pas été immergé dans une institution, j'ai travaillé chez moi ou dans les salles informatiques de l'école, j'avais également une liberté de choix totale quand à la solution informatique.
Ces conditions de travail ont – étonnamment ou non – plutôt représenté une difficulté supplémentaire à gérer.
- J'ai bien évidemment développé mes compétences dans le domaine de la recherche sur Internet.
- Les domaines dans lesquels mon apprentissage a été le plus riche sont la création des pages web dynamiques ainsi que le travail avec des bases de données accessibles par internet.

Et bien évidemment, la rédaction de ce mémoire, de part sa taille et son rôle, a également été une expérience formatrice.

Mon dernier mot est dédié au courageux lecteur arrivé jusqu'ici : Merci !

5 Bibliographie

La bibliographie est organisée autour des principaux domaines abordés tout au long de ce travail de diplôme soit:

- Création de site web
 - PHP & MySQL
- Evaluation de l'information présente sur Internet
- Internet – généralités
- Portails
- Recherche sur Internet
- Web et l'emploi

Liste des abréviations utilisées

p. = page
Vol. = volume
No. = numéro

Création de site web

FICHTER, Darlene, Delivering the Goods : intranet databases for small and larges projects,

IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 3, p. 88-90

CONGER, Sue, MASON, Richard O., Planning and designing effective web sites, Cambridge, Course technology, 1998, 184 p.

ROSENFELD, Louis, MORVILLE Peter, Information architecture for the World Wide Web, Beijing, O'Reilly, 1998, 202 p.

■ PHP & MySQL

ATKINSON, Leon, Programmation PHP, Paris, CampusPress, 1999, 436 p.

DELCROS, Armand, PHP3 faq, [En ligne]

<http://perso.cybercable.fr/adalcros/docs/php/php.html> (Page consultée le 7 mai 2000)

GILMORE, W.J., Beginning MySQL tutorial,

IN: DevShed.com – The open source web development site, [En ligne].

http://www.devshed.com/Server_Side/MySQL/Intro (Page consultée le 5 mai 2000)

HERLLIER Jean-Marc, MERIGOD, Philippe, Pages web dynamiques, Paris, CampusPress, 1999, 326 p.

LEPEZENNEC, Hervé, BRAND, Yan, PHP: enfin un langage pour le Web!,

IN: Flash informatique, la revue du Service informatique central de l'EPFL, [En ligne]

<http://www.epfl.ch/SIC/SA/publications/FI00/fi-1-00/1-00-page3.html> (Page consultée le 5 mai 2000)

Evaluation de l'information sur Internet

BIBEAU, Robert, Grille d'évaluation d'un site web, [En ligne]

<http://ntic.org/vitrine/veille/textes/BIBgrille.html> (Page consultée le 19 juin 2000)

Evaluation de l'information présente sur Internet,

IN: Sentiers d'accès et pistes de recherche d'informations scientifiques et techniques sur l'Internet !, INSA de Lyon, [En ligne]

<http://csidoc.insa-lyon.fr/sapristi/fristi36.html> (Page consultée le 19 juin 2000)

KIRK, E., Evaluating information found on the Internet, [En ligne]

<http://milton.mse.jhu.edu:8001/research/education/net.html> (Page consultée le 19 juin 2000)

SMITH, Alastair, Criteria for evaluation of Internet information resources, [En ligne]

<http://www.vuw.ac.nz/~agsmith/evaln/index.htm> (Page consultée le 19 juin 2000)

TILLMAN, Hope N., Evaluating quality on the net, [En ligne]
<http://www.hopetillman.com/findqual.html> (Page consultée le 19 juin 2000)

Internet – Généralités

CASTELLS, Manuel, La société en réseaux, Paris, Fayard, 1998, 613 p.

DERAMAIX, P., L'Internet : chaos culturel ou vecteur de connaissance ?, [En ligne]
<http://www.multimania.com/blerbe/internet.htm> (Page consultée le 10 juin 2000)

FLORIDI, Luciano, Brave.Net.World : the Internet as a disinformation superhighway ?,
IN: The electronic library, 1996, Vol. 14, No. 5, p. 509-514

JULLIARD, Laurence, Les internautes profilés,
IN: Webdo, [En ligne]
http://www.webdo.ch/webdo_mag/wm_03_2000/actu_baro_03.html (Page consultée le 3 juillet 2000)

TRASENSTER, Mark, New ways of delivering information,
IN: The electronic library, 1997, Vol. 15, No. 5, p. 395-397

Portails

BECKETT, Chris, Whose brand is it anyway ? Is there life after portal for content producers ?,
IN: Online information 99 : proceedings [of the] 23rd International Online Information Meeting, National Hall and Olympia 2, London, UK, 7-9 December 1999, Oxford, Learned Information Europe Ltd, 1999, p. 21-24

BLANCHERIE, Jean-Marc, Mais où sont les clés du Portail ?,
IN: Veille : le magazine professionnel de l'intelligence économique, 2000, No. 34, p. 20-22

DEBAUX, Guy, Les portails d'information d'entreprise,
IN: Veille : le magazine professionnel de l'intelligence économique, 2000, No. 34, p. 13-18

CANALLE, Cléo, Portail,
IN: Le monde interactif, [En ligne].
<http://www.lemonde.fr/article/0,2320,41622,00.HTML> (Page consultée le 11 avril 2000)

GUENTHER, Kim, Customized data delivery through web portals : a case study,
IN: Online : the leading magazine for information professionals, 1999, Vol. 23, No. 6, p. 51-56

LECLUYSE, Eric, La bataille des portails,
IN: Le monde interactif, [En ligne]. <http://www.lemonde.fr/article/0,2320,8732,00.html>
(Page consultée le 11 avril 2000)

MCLAGAN, Don, Portals: all change or all the same ?,

IN: Online information 99 : proceedings [of the] 23rd International Online Information Meeting, National Hall and Olympia 2, London, UK, 7-9 December 1999, Oxford, Learned Information Europe Ltd, 1999, p. 19-20

O'LEARY, Mick, Vortals on the rise,

IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 2, p. 79-80

WHITE, Martin, Portals: here today, gone tomorrow ?,

IN: Online information 99 : proceedings [of the] 23rd International Online Information Meeting, National Hall and Olympia 2, London, UK, 7-9 December 1999, Oxford, Learned Information Europe Ltd, 1999, p. 9-11

Recherche sur Internet

ANDRIEU, Olivier, Abondance : recherche d'information, référencement et promotion de sites web, [En ligne],
<http://www.abondance.com> (Page consultée le 12 avril 2000)

ANDRIEU, Olivier, Méthodes et outils de recherche sur l'Internet, Paris, Eyrolles, 1997, 235 p.

BATES, Mary Ellen, The Internet. Part of a professional searcher's toolkit,

IN: Online : the leading magazine for information professionals, 1997, Vol. 21, No. 1, p. 47-52

CYBION, Veille.com : la communauté de l'intelligence en ligne, [En ligne],
<http://www.veille.com/> (Page consultée le 12 avril 2000)

HOCK, Randolph, Web search engines : (more) features & commands,

IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 3, p. 17-26

LARDY, Jean-Pierre, RISI - Recherche d'information sur l'Internet, [En ligne],

<http://www.adbs.fr/adbs/sitespro/lardy/risi.htm> (Page consultée le 3 avril 2000)

LAWRENCE, Steve, LEE GILES C., Accessibility of information on the web, IN: Nature, 1999, vol. 400, p. 107-109

NOTESS, Greg, Search engine inconsistencies,

IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 2, p. 66-67

NOTESS, Greg, The never-ending quest : search engine relevance,

IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 3, p. 35-40

NOTESS, Greg, Search engine showdown, [En ligne],

<http://searchengineshowdown.com> (Page consulté le 15 avril 2000)

REVELLI, Carlo, Intelligence stratégique sur Internet : comment développer efficacement des activités de veille et de recherche sur les réseaux, Paris, Dunod, 1998, 212 p.

SHERMAN, Chris, Reference resources on the web,
IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 1, p.
52-56

SNOW, Bonnie, The Internet's hidden content and how to find it,
IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 3, p.
61-66

SHERMAN, Chris, The future revisited : what's new with web search,
IN: Online : the leading magazine for information professionals, 2000, Vol. 24, No. 3, p.
27-34

SULLIVAN, Danny, Search engine watch, [En ligne],
<http://www.searchenginewatch.com>, (Page consulté le 13 juillet 2000)

Web et l'emploi

DHOQUOIS, Anne, Le guide de l'emploi sur Internet, Paris, L'Etudiant, 1999., 185 p.

E-recruting, un outil qui révolutionne le recrutement, [En ligne],
<http://broadcast.nomad.ch/firsttuesday> (Page consultée le 19 juin 2000)

JONES, V. C., A basic guide to job search resources on the Internet : career planning and
job searching in the information age,
IN: Reference librarian, 1996, No. 55

LEBAUBE, Alain, Les initiatives emploi : dossier,
IN: Le monde, 16 septembre 1998, p. 1

VAN EECKHOUT, Laetitia, Recruter sur Internet : les premiers pas timides des entreprises
françaises,
IN: Le monde, 19 janvier 1999, p. 6

6 Lexique

CGI : [acronyme utilisé pour] Common Gateway Interface.

Le CGI est une méthode de transfert, pour permettre à d'un ordinateur client de transmettre des données vers un serveur web (par exemple lorsque l'on envoie des informations à partir d'une page web, grâce à un formulaire). Le cgi n'est que le mode de transmission, les données sont ensuite traitées par un programme sur le serveur, qui renvoie ensuite une réponse adéquate au client

BRUIT: Information non-pertinente

EXTRANET : Un extranet est un réseau similaire à un intranet, mais accessible depuis n'importe où via Internet et grâce à un mot de passe.

HTML : [acronyme utilisé pour] Hyper Text Mark Up Language Norme utilisée pour baliser et structurer un document, permettant ainsi sa publication sur le web.

INTRANET : Un intranet est une sorte d'Internet propre à une entreprise ou à une institution, un intranet est composé de pages et de services qui utilisent les mêmes protocoles que ceux que l'on trouve sur Internet. L'intranet d'une entreprise ou organisation est uniquement consultable à partir du réseau local de celle-ci.

META-TAGS : Les meta-tags sont une méthode de description du contenu d'un document HTML. L'idée étant d'inclure une sorte de description bibliographique au sein même du document – de l'information sur l'information. Les meta-tags peuvent contenir différents types d'information (auteur du document, mots-clés en décrivant le contenu etc...). Ces meta-tags sont invisibles lors du visionnement de la page HTML, ils sont cachés dans une section du document réservée à cet effet.

NEWSGROUP : Forum de discussion. Dans un newsgroup les messages (envoyés par mail) sont pris en charge par un réseau de serveurs qui se charge de mettre ceux-ci à disposition. Chaque utilisateur peut ensuite consulter ce réseau de serveur pour consulter les nouveaux messages du newsgroup. Ce système s'oppose à celui des listes de discussion, où chaque participant reçoit automatiquement l'intégralité des messages de la liste par courrier électronique.

PUSH : Système de d'information sur Internet, au lieu de rechercher activement l'information, les logiciels push offrent un mécanisme d'abonnement aux utilisateurs, par ce biais, l'internaute sélectionne des thèmes l'intéressant, et le logiciel push, lui fait régulièrement parvenir des informations relatives aux thèmes qu'il aura sélectionnés.

URL : [acronyme utilisé pour] Unified Ressource Locator

C'est ainsi que sont appelées les adresses des sites web, pour plus de clarté voici un exemple choisi au hasard : <http://www.unige.ch/esid>

USENET : C'est ainsi qu'est appelé le réseau de serveur relayant les messages des newsgroups.

VEILLE STRATEGIQUE : Activité d'une entreprise ou institution ayant pour but de collecter et de diffuser les informations nécessaires à la prise de décisions stratégiques.

WYSIWYG : [acronyme utilisé pour] What You See Is What You Get,

Les éditeurs HTML permettant de travailler directement de manière graphique sur un document sont qualifiés d'éditeurs WYSIWYG. Ce que l'on voit à l'écran sera notre résultat final. En appliquant cette définition à un autre domaine, l'on pourrait qualifier le logiciel Word d'éditeur WYSIWYG pour feuille A4.