



**INSTITUT NATIONAL DES LANGUES ET CIVILISATIONS
ORIENTALES**

Centre de Recherche en Ingénierie Multilingue

Mémoire de DESS Ingénierie Multilingue

**NTIC & ÉTHIQUES... QUELLE VALEUR
POSSÈDE L'INFORMATION EN LIGNE?**

**Enjeux liés à l'information et conséquences de la rencontre entre
Marketing et information en ligne**

Jean –Louis GARÇON

Sous la direction de :

Claude de Loupy & Jean-Michel Daube

ANNÉE 2004-2005

*** Remerciements ***

Je tiens à remercier l'ensemble de l'équipe du Centre de Recherche en Ingénierie Multilingue de l'INALCO pour l'année enrichissante que j'ai passée à leurs côtés, d'un point de vue académique, autant que d'un point de vue humain.

Je remercie mes camarades du DESS Ingénierie Multilingue et du DESS Traductique et Gestion de l'Information de l'INALCO, pour leur gentillesse et leur grande ouverture d'esprit car j'ai également progressé grâce à eux.

Je remercie enfin plus particulièrement mes directeurs de mémoire, Monsieur Claude De Loupy, et Monsieur Jean-Michel Daube, pour leurs conseils, leur soutien dans les moments difficiles qui ont ponctué la rédaction de ce mémoire, et leur disponibilité.

Merci !

Résumé

L'avènement de la société de l'information redéfinit l'économie en une « nouvelle économie », numérique, désintermédiée, réintermédiée, dans laquelle l'entreprise doit apprendre à communiquer différemment pour profiter des opportunités commerciales que lui offre le Web.

Le consommateur quant à lui, ne se contente plus de subir l'information en spectateur passif, il navigue sur le Web, il cherche, il trouve, il compare et il échange les informations qu'il rencontre. C'est une toute nouvelle donne...

De spectateur, il devient acteur de la nouvelle économie en diffusant sa propre information, et ce, volontairement ou non.

Le rêve de la bibliothèque d'Alexandrie n'est pas si loin, mais le problème est que le Web n'a rien d'une bibliothèque dans le fonctionnement du référencement de ses contenus et de leur validation.

Les outils de recherches – moteurs de recherche en tête – ont remplacé les documentalistes et sont aujourd'hui plébiscités par les internautes et par les entreprises lors d'une recherche d'information.

En effet, chaque partie tire profit de la simplicité d'utilisation des moteurs ou des modèles économiques qui y sont liés, avec pour conséquence principale, l'introduction de biais supplémentaires, dans la recherche et dans le traitement de l'information par le consommateur final, ce dernier devant prendre certaines précautions afin de valoriser l'information récoltée.

Mots-clés

Moteur de recherche, information, e-business, vente de mots-clés, modèle d'affaires, méthodologie de recherche, valeur, multilinguisme, NTIC.

Intitulé et adresse de l'UFR

Institut National des Langues et Civilisations Orientales - C.R.I.M
Centre de Recherche en Ingénierie Multilingue
2, rue de Lille
75343 Paris Cedex 07

Abstract

The Advent of an information society reshapes our traditional economy in a « new economy », a digital economy, disintermediated, reintermediated, and where every single actor must learn how to communicate differently, in order to benefit from the new commercial promises of the Web.

The consumer no longer undergoes the information as a docile viewer, but is now able to surf the Web, look for information, find it, compare it, and swap it with the Internet population. It's a whole new game.

From witness, the consumer becomes an actor in the new economy, spreading its own information, willingly or not.

The dream of the Alexandria library is not that far away, but the problem is that the Web has nothing to do with a library in the way it operates indexing and content authentication.

Search tools – especially search engines – have replaced librarian and they are now used by an overwhelming majority of the Internet population and of businesses, too.

Each entity finds its advantage in the simplicity of use and in the business models which are related to search engines.

This leads to the interference of new biases in the field of information retrieval, as the final user needs to be very careful in order to create value from the information he or she has collected.

Keywords

Search engine, information, e-business, contracts on words, business model, research methods, value, multilinguism, IT.

TABLE DES MATIÈRES

* <i>Remerciements</i> *	3
Résumé.....	4
Mots-clés.....	4
Intitulé et adresse de l’UFR	4
Abstract.....	5
Keywords	5
INTRODUCTION	9
I. INTRODUCTION	10
I. 1Le cadre :	10
I.2 Qu’est-ce qu’une information, une connaissance, un document?	12
Définition de Sylvie Fayet-Scribe.....	12
L’explosion de l’information à l’ère du numérique.....	14
II. L’explosion de l’information à l’ère du numérique.....	15
II.1 TIC,NTIC et Société de l’information... ..	15
II.1.a Tic & Ntic.....	15
II.1.b Société de l’information	16
II.2 Dissection du Web	17
II.2.a Structure et propriétés.....	17
II.2.b Web visible et Web invisible	20
II.3 Production et recherche d’information.....	21
II.4 La diversité du contenu présent sur le Web	22
II.4.a La construction du Web et de son industrie	22
II.4.b La promesse commerciale du Web.	23
II.4.c Un marché... potentiel	24
II.4.d L’équipement des ménages	24
II.4.e L’équipement des entreprises	25
II.5 L’autre impact des sites commerciaux	26
II.5.a Acheter sur le Web	27
II.5.b Les principaux freins à l’achat sur Internet.....	27
Naissance des outils de recherche et fonctionnement.....	29
III. Naissance des outils de recherche et fonctionnement.....	30
III.1 Des problématiques croisées.....	30
III.2 Définitions des notions de base.....	31
III.2.a Qu’est-ce qu’un moteur de recherche ?	31
III.2.b Qu’est-ce qu’un portail ?	32
III.2.c Qu’est-ce qu’un métamoteur ?	32
III.2.d Qu’est-ce qu’un annuaire... ..	33
III.2.e Qu’est-ce que le Text-Mining ?	34
III.3 Comment fonctionnent les moteurs de recherche ?.....	34
III.3.a Premier composant : L’interface graphique	34
III.3. b Deuxième composant : L’index et l’opération d’indexation	35
III.3.c Troisième composant : Le module d’appariement.....	40
III.3.d Quatrième composant : La base documentaire	41
III.3.e Cinquième composant : L’utilisateur	42
III.4 La composition des moteurs, en résumé :.....	42
III.5 État de l’art de la technique en matière de recherche d’information	42
III.5.a Approche générale	43

III.5.b Google.....	44
III.5.c Yahoo !.....	46
Pourquoi les outils de recherche ont-ils tant d'intérêt pour le commerce électronique ?	47
IV. Pourquoi les outils de recherche ont-ils tant d'intérêt pour le commerce électronique ?.....	48
IV.1 Le rôle des outils de recherche	48
IV.1.a Comment les internautes utilisent-ils les outils de recherche ?	48
IV.1.b Attrait en terme de trafic généré et de communication.....	49
IV.1.c Attrait en terme d'accès à l'information	50
IV.2 Une nouvelle manière d'échanger de l'information	50
IV.2.a La spécialisation des moteurs	51
IV.2.b La personnalisation des moteurs et les réseaux sociaux.....	52
IV.3 Le Web et la culture du gratuit	53
Modèles d'affaires des moteurs de recherche	54
V. Modèles d'affaires des moteurs de recherche.....	55
V.1 Le Keyword Marketing ou le commerce des mots-clés.....	56
V.1.a L'innovation du coût par clic.	56
V.1.b La soumission payante ou « insertion payante ».....	58
V.2 Les méthodes de référencement alternatives, abusives:.....	59
V.2.a Le détournement de l'utilisation de la balise <Title>	59
V.2.b L'utilisation d'informations invisibles (« Le texte invisible »).....	59
V.2.c Le cloaking (ou technique des pages masquées).....	60
V.2.d Les liens hypertextes dissimulés	60
V.2.e L'utilisation de redirections trompeuses	60
V.2.f L'utilisation de liens au sein d'un réseau (le « Google Bombing »)	60
V.2.g La fraude au clic.....	61
Conséquences de ces business models sur le monde de l'information de la connaissance et des langues	63
VI. Conséquences de ces business models sur le monde de l'information de la connaissance et des langues	64
VI.1 Le problème de la pertinence de l'information sur Internet	64
VI.2 État de conscience des consommateurs (utilisateurs des moteurs de recherche)	66
VI.3 Les informations personnelles/ La CNIL	68
VI.3.a La CNIL	68
VI.3.b Nos informations personnelles	71
VI.4 Comment adapter le modèle et protéger le consommateur?.....	76
VI.5 L'importance d'une démarche pédagogique :	78
VI.5.a Le constat:	78
VI.5.b Les types de question auxquelles l'utilisateur peut être confronté:	78
VI.5.c Étapes clés de la recherche d'information :	79
VI.5.d Etablir une stratégie de recherche :.....	80
VI.5.e Les conseils pour une meilleure recherche d'information sur le Web.....	80
VI.6 Web, langues et information.....	82
VI.6.a Qu'est ce donc qu'une langue ?.....	83
VI.6.b Permettre au Web de devenir multilingue	84
Les moteurs de recherche et le droit	92
VII Les moteurs de recherche et le droit.....	93
VII.1 Le positionnement payant face aux règles du droit de la concurrence.....	94

VII.2	Le positionnement payant : quelle protection pour les consommateurs ?.....	95
VII.3	L'interdiction de la publicité clandestine ou trompeuse.	96
VII.4	L'application de la loi Sapin aux intermédiaires de positionnement payant..	97
VII.5	Le “ position squatting ” et la protection des marques.....	97
VII.5.a	La responsabilité du squatteur	98
VII.5.b	La responsabilité de l'outil de recherche.....	101
VII.5.c	La responsabilité civile délictuelle du prestataire de positionnement payant.....	102
VII.5.d	La responsabilité pénale du prestataire de positionnement payant.	103
VII.6	Le droit d'auteur.....	104
VII.6.a	Difficultés d'application des règles juridiques sur Internet.....	105
VII.6.b	Les difficultés d'application de la législation liées à la dimension internationale d'Internet.....	108
VII.6.c	Le réseau doit s'adapter au droit.....	112
VII.6.d	Le droit doit s'adapter au réseau	113
VII.6.e	La nécessité d'une entente internationale.....	114
VII.7	La coopération internationale.....	115
VII.7.a	Au niveau européen	115
VII.7.b	Au niveau mondial	116
	Les autres sources d'information sur le Web.....	117
VIII.	Les autres sources d'information sur le Web.....	118
VIII.1	Blogs (joub, Weblogs, Webillards).....	118
VIII.2	Les banques de données.....	118
VIII.3	Forums (agora, ou newsgroups).....	119
VIII.4	Annuaire spécialisés	120
VIII.5	Presse en ligne.....	120
VIII.6	Services de puges (agents, RSS).....	122
VIII.7	La baladodiffusion ou « podcasting »	122
VIII.8	Les encyclopédies collaboratives en ligne	123
VIII.9	Le projet Google Print.....	124
	Les Enjeux liés à l'information.....	126
IX.	Les Enjeux liés à l'information.....	127
IX.1	Tous égaux devant l'information ?.....	127
IX.1.a	Corée du Nord et Afghanistan	127
IX.1.b	Chine.....	127
IX.1.c	France et États-Unis.....	128
IX.2	A qui revient la responsabilité de contrôler l'information Web ?	129
IX.3	Comment construire une charte de bonnes pratiques de l'Internet ?.....	131
IX.4	D'importants enjeux sont liés au contrôle et à la diffusion de l'information.....	132
	Conclusion et perspectives.....	135
X.	Conclusion et perspectives.....	136
	Bibliographie.....	139
-	Bibliographie.....	140
	Articles/Rapports/Enquêtes/ Études :	141
	Webographie	144
	ANNEXES.....	147
-	ANNEXES.....	148

INTRODUCTION

Parce qu'il y a un début à tout...

I. INTRODUCTION

I. 1 Le cadre :

Bienvenue dans un monde numérique : nous avons quitté la grande société industrielle et entamons l'ère de la société de l'information : croissance exponentielle de l'informatique au sein des ménages, baisse des prix de l'équipement, facilité d'accès aux connections haut débit, besoins en information croissants tant dans la sphère privée que professionnelle, et explosion du nombre de données textuelles accessibles sur la Toile.

Cet avènement de la société de l'information redéfinit l'économie en une « nouvelle économie », numérique, désintermédiée, réintermédiée, dans laquelle l'entreprise doit apprendre à communiquer différemment pour profiter des opportunités commerciales que lui offre le Web.

Le consommateur quant à lui, ne se contente plus de subir l'information en spectateur passif, il navigue sur Internet, il cherche, il trouve, il compare et il échange les informations qu'il rencontre. C'est une toute nouvelle donne...
De spectateur, il devient acteur de la nouvelle économie en diffusant sa propre information, et ce, volontairement ou non.

Le rêve de la bibliothèque d'Alexandrie n'est pas si loin, mais le problème est que le Web n'a rien d'une bibliothèque dans le fonctionnement du référencement de ses contenus et de leur validation.

Les ouvrages que nous trouvons sur les étagères des bibliothèques ont été classés et référencés par des professionnels, humains, ayant reçu une formation, et dotés d'un esprit critique et de logique. Ils reconnaissent les auteurs des ouvrages pour leurs travaux, et leurs œuvres sont regroupées par familles, catégorie ; elles s'accompagnent de notices, de commentaires, etc.

Dans une bibliothèque l'information est validée en amont, tous les efforts sont mis en œuvre pour que l'utilisateur du document puisse avoir une confiance optimale en l'information « pré-parée » qu'il reçoit...qu'en est-il sur le Web ?

De plus, la rencontre entre information et commerce sur le Web introduit certains biais supplémentaires dans le traitement de l'information, et impose davantage de précautions pour le consommateur final.

L'internaute se trouve donc constamment face à un risque dès lors qu'il recherche de l'information sur le Web, sans qu'il en soit forcément conscient.

Ceci dit, le Web représente aussi un formidable espace de liberté, un espace d'expression et d'échanges (quels qu'ils soient), où chacun a la possibilité de produire et de diffuser de l'information.

C'est pour cela que l'information croît sur le Web à une vitesse exponentielle et qu'il est délicat de gérer cette masse de contenus.

De nouveaux services et outils, censés faciliter la tâche des utilisateurs, ont donc progressivement fait leur apparition, comme les navigateurs Internet (Web browsers¹) qui ont offert aux internautes la possibilité de naviguer (« surfer ») d'une page Web à une autre en soumettant manuellement l'adresse de chaque page à afficher. Mais le véritable besoin de la communauté Internet deviendra rapidement celui de disposer d'un outil capable de rendre la recherche d'information plus facile, plus intuitive, de trouver une page automatiquement, même lorsque l'on ne connaît pas son adresse.

La réponse à cette attente est arrivée vers le milieu des années 1990, avec les annuaires puis avec un outil qui va rapidement évoluer pour devenir un acteur incontournable du Web : le moteur de recherche.

Le moteur de recherche est aujourd'hui plébiscité par les internautes et par les entreprises lors d'une recherche d'information, chaque partie tirant profit de la simplicité d'utilisation ou des modèles économiques qui y sont liés.

Mais qu'est-ce au juste qu'un moteur de recherche? Comment cela fonctionne-t-il? Comment cet outil a-t-il réussi à s'imposer auprès des internautes jusqu'à devenir leur principal point d'entrée sur le Web? Pourquoi est-il devenu une donnée indispensable dans la stratégie commerciale de toute entité souhaitant développer sa visibilité sur le Web? Comment sont financés les moteurs? Ces modèles économiques ont-ils des conséquences sur le monde de l'information, de la connaissance et des langues? Que représente l'information en terme d'enjeux économiques, politiques et sociaux, et quelles responsabilités avons-nous lorsque nous utilisons cette information en libre service? L'information gratuite peut-elle être de qualité?

Comment chercher de l'information pour créer de la valeur? Voici quelques-unes des questions auxquelles cette étude s'efforcera de répondre.

¹ Par ex. Internet Explorer ou Netscape Navigator, ou encore Mozilla Firefox

1.2 Qu'est-ce qu'une information, une connaissance, un document?

Définition de Sylvie Fayet-Scribe²

« Il faut tout d'abord poser les différentes définitions car il y a un problème de polysémie des termes se rattachant aux divers concepts d'information, de document, de connaissance :

- "**L'information** spécialisée est une information destinée aux spécialistes (industriels, chercheurs, enseignants) qui leur est nécessaire dans leur activité professionnelle et qui conditionne une prise de décision ou une action particulière. Il faut la différencier de l'information pour le grand public notamment par la nature des informations transmises ainsi que le canal de communication emprunté."³

Historiquement, la définition est plus large car l'information, plus globalement est liée à la connaissance, et à des formes de connaissance diverses : connaissance théologique, mythique, philosophique, technique, scientifique... que vont s'approprier divers individus qui n'ont pas encore de statut professionnel dans la société, tout en étant liées à des lieux privilégiés de la connaissance : bibliothèques, monastères, corporations, universités, laboratoires, académies, entreprises... Peu à peu se formera "une information scientifique et technique" (IST), de portée internationale, indépendante d'une culture particulière et validée par l'ensemble d'une communauté. A ce titre, les prémices de l'IST sont sans doute à voir au XVIIe siècle, moment où naissent tout à la fois les périodiques scientifiques, les laboratoires et des individus qui ont pour objectif l'étude scientifique et/ou technique; de plus, la communauté internationale scientifique a une activité sans frontière.

Quant au mot information, au sens étymologique du mot, "in-former" signifie "donner une forme", or la matière première à mettre en forme est bien **la connaissance** :

- **Connaître** signifie à la fois savoir et comprendre, c'est-à-dire être capable d'expliquer le rapport de l'objet de connaissance avec les autres choses déjà sues.

Cette mise en forme de la connaissance se fera d'abord par le langage, puis par l'écriture, et celle-ci s'inscrira dans **le document** :

- Le document écrit imprimé (mais aussi le document électronique, photographique, vidéographique... ou autre) peut être défini selon l'indication de l'Organisation Internationale de Normalisation (ISO-TC 46) comme un : "Ensemble formé par un support et une information, généralement enregistrée de façon permanente et tel qu'il puisse être lu par l'homme ou la machine"; Ce que traduit bien la définition proposée par J. Meyriat et R. Estival : "Toute connaissance mémorisée, stockée sur un support, fixée par l'écriture ou inscrite par un moyen mécanique, physique, chimique, électronique, constitue un document. Dans la chaîne primaire de création, tant que le document est créé

² Sylvie Fayet-Scribe est maître de conférences à l'Université Paris-I Panthéon-Sorbonne, où elle est responsable de la maîtrise de sciences de l'information et de la documentation.

³ •Chartron, 1992, p.351.

mais n'est pas encore utilisé par un récepteur, il reste un document virtuel. Dès lors qu'il est employé et décodé par un récepteur, le document devient réel : il a trouvé son utilisateur"⁴.

Utiliser un document, y puiser une information est un acte fondamental. Nous avons pour objectif d'étudier les voies d'accès à [cette] information, via le document et son support, à travers des outils qui sont capables de fournir les références du document, voire les informations elles-mêmes. ». Nous étudierons également les conditions de production de l'information, en commençant par présenter le contexte dans lequel nous évoluons actuellement, celui d'une société de l'information, organisée autour de l'utilisation des nouvelles technologies.

⁴ Meyriat et Estival, 1981, p.84

L'explosion de l'information à l'ère du numérique

Cette partie présente l'avènement d'une nouvelle société : la société de l'information, caractérisée par le développement des Nouvelles Technologies de l'Information et de la Communication (NTIC), et l'essor du Web. Nous rappellerons la distinction qu'il y a lieu de faire entre web et internet. Nous examinerons le Web plus en détails afin d'en dégager la structure, et de présenter sa composition et ses potentialités en terme de contenus. Nous constaterons également l'abondance d'informations disponibles sur le Web et nous verrons quelles en sont les conséquences en prenant notamment en compte les effets liés à l'arrivée du monde des affaires sur le Web.

Cette partie présente également l'apparition des premiers outils du Web et des premières offres commerciales (Fournisseurs d'Accès Internet en tête). Nous traiterons de la recherche de visibilité globale liée à la promesse commerciale faite par le Web aux entreprises et aux particuliers : un nouvel espace potentiel d'échange(s) qui suppose un équipement minimum.

Nous verrons de quelle manière cette cohabitation modifie les modes de consommation et de communication pour les consommateurs et pour les sites commerciaux. Enfin, nous montrerons de quelle manière les problématiques croisées des différents acteurs du Web et la quantité d'information disponible feront naître le besoin d'outils permettant de faciliter l'accès aux (res)ources...

II. L'explosion de l'information à l'ère du numérique

II.1 TIC, NTIC et Société de l'information...

II.1.a Tic & Ntic

Les TIC (les Technologies de l'Information et de la Communication⁵) sont l'ensemble des technologies qui permettent la transmission en temps réel des données numériques⁶.

Les TIC regroupent un ensemble d'appareils nécessaires pour manipuler de l'information, (en particulier des ordinateurs) et les programmes nécessaires pour convertir cette information, la stocker, la gérer, la transmettre et la retrouver.

Les premiers pas vers une société de l'Information furent entamés lors de l'invention du télégraphe électrique⁷, du téléphone fixe⁸, de la radiotéléphonie⁹ et, enfin, de la télévision¹⁰. La télécommunication mobile (1979), l'Internet (1983), et le GPS (1995) ; le Blackberry (1999) peuvent être considérés comme des NTIC (Nouvelles Technologies de l'Information et de la Communication).

⁵ en Anglais : IT (Information Technology) ou ITC (Information and Communication(s) Technology). Le « s » de communications est sujet à controverse.

⁶ Source thèse de Master of Sciences du CIHEAM-n°59, p.12

⁷ en 1838 par Charles Wheatstone

⁸ Le téléphone a été inventé dans les années 1860 par Antonio Meucci, un italien émigré aux États-Unis (ce fait a été récemment reconnu par le congrès américain dans la résolution 269 du 15 juin 2002). Avant cette résolution, l'invention du téléphone était attribuée à Alexander Graham Bell qui déposa un brevet en 1876. Meucci avait fait une demande de brevet en 1871 mais faute de ressources, elle expira en 1874[Wikipédia]

⁹ 1896, premier brevet déposé par Marconi

¹⁰ 1860, première transmission d'images (fixes) par voie électrique effectuée par Caselli avec un appareil baptisé pantélégraphe.

II.1.b Société de l'information

La notion de société de l'Information a été introduite et popularisée en Europe en 1994 avec la publication du rapport Bangemann intitulé «Europe and the global information society». Identique aux concepts de NII¹¹ et GII¹² mais avec un intérêt moindre pour les technologies, au profit des utilisations.

La spécificité de la société de l'information est la position centrale qu'y occupent les technologies de l'information et de la communication (cf. chapitre dédié) en matière de production et d'économie. L'Information devient le cœur du système, elle est essentielle à l'existence des autres produits. Le succès des entreprises et des organisations repose sur la capacité de ces entités à exploiter l'Information, et la plupart des salariés dépendent des flux d'information pour accomplir leur travail. En pratique, cette Information dépend directement des nouvelles technologies et de l'Internet.

Des ordinateurs de plus en plus performants, un besoin d'être informés de plus en plus important, de plus en plus d'information sur la Toile (cf. Taille du Web) et dans les bases de données des entreprises, de plus en plus de supports d'information différents, voilà bien les contours de la société dans laquelle nous évoluons aujourd'hui.

On considère que cette société de l'information succède à la société industrielle.

Les concepts suivants sont tous intimement liés à cette notion :

- ⇒ « société post-industrielle » (Daniel Bell),
- ⇒ « post-fordisme »,
- ⇒ « société post-moderne »,
- ⇒ « société de la Connaissance »,
- ⇒ « société Télématique »,
- ⇒ « Révolution de l'Information »,
- ⇒ « société informationnelle ».

Les intérêts qui sont à l'heure actuelle en jeu sur le Web, sont de nature stratégique. De même que la société industrielle consistait à s'approprier les ressources énergétiques pour assurer son essor, le pouvoir des individus, des entreprises, et des États, est aujourd'hui fonction de leur accès à l'information.

¹¹ National Information Infrastructure (cf. <http://www.christlinks.com/glossary2.html>)

¹² Global Information Infrastructure (cf. <http://www.christlinks.com/glossary2.html>)

II.2 Dissection du Web

Il est important de rappeler ici la différence entre l'Internet (ou « le Net »), et le Web (www.) car une confusion demeure aujourd'hui encore dans l'esprit de bon nombre de personnes :

- **L'Internet**, désigne le réseau **physique** constitué de machines reliées les unes aux autres par des câbles, ou des ondes pour les réseaux sans fil.

Je ne citerais pas ici de données concernant le nombre de machines connectées au réseau car ces chiffres semblent par nature très discutables, d'autant plus qu'avec l'explosion des connections sans fil, l'évaluation se complique.

- **Le Web**, désigne quant à lui l'espace **virtuel** qui s'étend entre les bornes du réseau physique ; c'est cet espace que nous visualisons sur nos écrans (à l'aide de navigateurs comme Internet Explorer, Mozilla Firefox, etc., nous permettant d'afficher des pages se trouvant sur les machines connectées au réseau.).

II.2.a Structure et propriétés

La structure du Web est fondée sur le principe des hyperliens qui sont autant de routes entre les pages. Toutes les pages Web ont une adresse unique appelée URL (Unified Resource Locator). L'URL est une suite de caractères (sans espace) qui définit l'endroit (locator) où se trouve un document (resource).

L'URL est l'adresse standard de n'importe quelle page ou document, sur n'importe quel serveur connecté sur Internet.

Elle se compose :

- du protocole de communication
- du nom du serveur (domain name) où se trouve le document
- du nom du répertoire (éventuellement) et du nom de la page

Et elle s'écrit :

Protocole://serveur/répertoire/page.extension

Exemple d'adresse URL d'une page de l'INALCO traitant de l'extranet :

<https://www.inalco.fr/extranet/>

Le http (Hyper Text Transfert Protocol) est le protocole de communication, entre un navigateur (le client) et un serveur connecté sur Internet, permettant le transfert des pages Web. Dans l'exemple ci-dessus le https correspond à la version sécurisée (cryptage des données) du protocole http.

Il existe d'autres protocoles sur Internet comme le FTP (File Transfert Protocol) qui permet de télécharger des fichiers à distance, ou comme le Mailto pour le courrier électronique. Le langage permettant de décrire les pages Web est le HTML (Hyper Text Markup Language). Ce langage permet de doter certains mots ou certaines images, appelées des ancres, de la fonction d'hyperlien. Cette fonction permet, par un simple clic de souris, de passer directement d'une partie d'un texte à une autre ou d'une page à une autre n'importe où sur Internet.

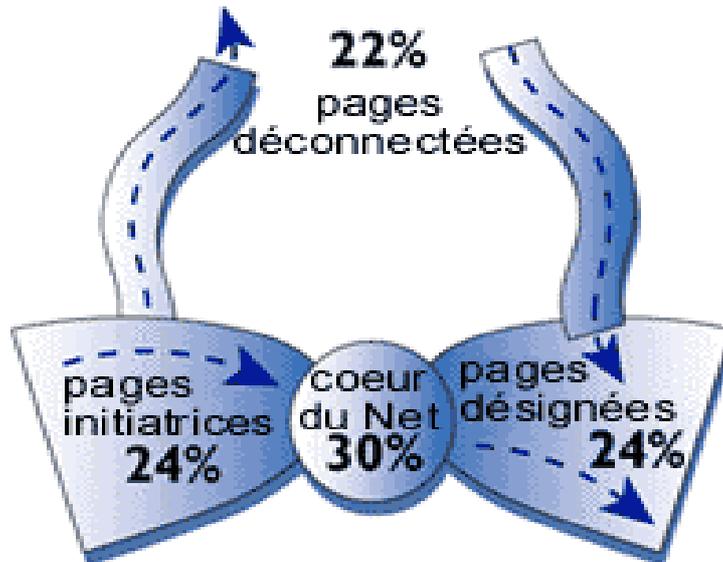
Le World Wide Web serait donc une « Toile d'araignée » (d'où son nom) de serveurs d'informations, reliés les uns aux autres par des liens physiques (le réseau matériel) et des liens logiques, intangibles (les hyperliens).

Cette toile d'araignée, par son immensité et sa progression constante tant en nombre de nœuds (les pages Web) que de liens (les hyperliens), est souvent perçue comme étant infinie et sans réelle structure.

Cependant, une équipe de chercheurs d'AltaVista, de Compaq, et d'IBM a établi une cartographie du Web. Contrairement à l'image consacrée et citée précédemment, Internet n'aurait pas la forme « d'une toile d'araignée », mais plutôt celle d'un « nœud papillon ».

L'étude réalisée entre mai et octobre 1999, par les chercheurs d'AltaVista, de Compaq, et d'IBM, sur près de 500 millions de pages Web, a permis de construire le schéma suivant décrivant la connectivité d'Internet.

Le Web et la théorie du « nœud papillon ».



Le Web serait ainsi divisé en quatre grandes régions :

Le cœur du Net : 30% des pages Web constituent de véritables carrefours. De nombreux liens y mènent et ces sites renvoient sur de nombreux autres sites.

Pages déconnectées : 11% des pages ne sont reliées à aucune autre. Pour y accéder, il faut impérativement en connaître l'adresse. 11% des pages s'échappent des pages initiatrices pour aboutir à des culs-de-sac.

Pages initiatrices : 24% des pages Web renvoient vers des sites appartenant au cœur du Net, sans pour autant que la réciproque soit vraie. Il s'agit probablement de nouveaux sites qui n'ont pas encore été découverts, et donc pas encore reliés à ceux du cœur.

Pages désignées : 24% de ces pages sont fréquemment désignées, mais, en revanche, ne renvoient sur aucun site.

Cette théorie du nœud papillon a permis la découverte de frontières rendant la navigation difficile voire, dans certains cas, impossibles entre certaines régions du Web.

Cela nous donne un aperçu du monde dans lequel les outils de recherche ont dû évoluer pour se doter de stratégies leur permettant de rendre accessible la plus grande partie du réseau possible¹³.

¹³ Rakovec, 2002 : http://cj.tronquet.free.fr/cnam/pr_detail.php?dl=54

II.2.b Web visible et Web invisible

Le Web se compose du **Web visible** (Surface Web) et du **Web invisible**¹⁴ (“deep Web” ou encore “hidden Web”).

1. Le « **Web visible** » désigne l’ensemble de l’information disponible sur le World Wide Web pouvant être retrouvée à l’aide des moteurs de recherche. On considère que dès lors qu’une information du Web invisible peut être retrouvée par un moteur, elle appartient au Web visible.

Sa taille est estimée à 11,5 milliards de pages [GULLI & SIGNORINI, 2005].

Mais toutes ces pages ne sont pas informatives, notamment en raison des dérives liées au référencement des sites Web sur les moteurs qui entraînent une floraison de pages satellites et autres doublons (forme de spamdexing¹⁵).

Une petite partie des informations est indexée (environ 4 milliards de pages) :

- Celles qui sont facilement accessibles aux outils,
- Celles qui font la démarche pour se faire connaître des outils
- Beaucoup de sites commerciaux mais aussi
- Des sites émanant d’organismes, d’institutions (bibliothèques, universités, organismes nationaux et internationaux,...) ou d’états (ministères)
- Des sites personnels.

Ces sites sont eux même regroupés en « sites actifs », ceux sur lesquels nous naviguons, en « sites en construction », dont le contenu n’est pas exploitable, et en sites non pertinents, (doublons, copies de sites actifs, pages contenant des erreurs)

2. Le « **Web invisible** » désigne l’ensemble des pages publiquement accessibles sur le World Wide Web, mais qui ne sont pas accessibles par des liens hypertexte, c’est-à-dire qu’on ne peut y accéder qu’en saisissant directement leur URL (du type <http://www.quelquechose.com>), dans la barre d’adresse d’un navigateur. Le Web Invisible est principalement constitué de pages Web générées de manière automatique via des requêtes effectuées auprès de bases de données (pages protégées par un mot de passe, archives, pages personnelles créées pour un nombre limité d’internautes).

Le Web invisible, c’est ce que les outils de recherche ne voient pas :

Les outils n’indexent que le début des pages;

Ils n’indexent que du texte (les noms et informations « métadonnées » sur les images et les sons)

Ils n’indexent que les premiers niveaux des sites.

Ils n’indexent que ce qui n’est pas protégé par mots de passe (privé ou payant)

Ils n’indexent pas (ou mal) les pages dynamiques des bases de données

Ils n’indexent pas tous les types de fichiers (.pdf, .doc, etc.)

¹⁴ A ne pas confondre avec le “Dark Web” ou “Dark Internet” qui désigne pour sa part, l’ensemble des machines qui ne sont pas accessibles par d’autres ordinateurs sur l’Internet.

¹⁵ Techniques de fraude à l’intérieur du code HTML d’une page pour tenter de tromper les algorithmes de tri des moteurs de recherche. Exemple : répéter 100 fois le même mot, ou le nom des concurrents, insérer du texte en blanc sur fond blanc, répéter la balise de titre, etc.

Le spamdexing est un néologisme anglais composé du substantif spam et du suffixe -dexing pris sur le terme indexing signifiant référencement. En français, on peut traduire spamdexing par référencement abusif. L’objet du spamdexing est de tromper un catalogue ou un système de classement électronique, pour augmenter la fréquence de citation du site dans les résultats de recherche.

La seule façon d'accéder à un tel site, c'est de connaître son adresse précise et les éventuels mots de passe.

La taille du Web invisible a été estimée à 550 milliards de pages en 2001 [BERGMAN, 2001].

Certains outils permettent de chercher dans l'invisible comme le métamoteur <http://www.profusion.com> ou les répertoires sélectifs <http://www.invisible-Web.net> et <http://www.completeplanet.com> (70.000 bd) ou encore <http://urfist.uni-lyon1.fr>. Il existe également les archives ouvertes <http://www.ccsd.cnrs.fr> et <http://arxiv.org> et enfin la « machine à remonter le temps¹⁶ » <http://www.archive.org>.

A titre d'exemple, Google affiche sur la page d'accueil de son site un index d'un peu plus de 8 milliards de pages Web sur les 562 milliards que l'on attribue au Web (visible + invisible).

II.3 Production et recherche d'information

Ce n'est un secret pour personne : la quantité d'information disponible sur Internet est vertigineuse. Cela est dû à une propriété intrinsèque de l'information : la démultiplication.

Une information, lorsqu'elle est échangée, créée, comparée, engendre de nouvelles informations. Propriété qui se voit exacerbée par la nature « intégrante » des TIC.

En effet, comme le dit très justement Joël de Rosnay, si l'automobile remplace la charrette tirée par des chevaux (principe de substitution), « l'écrit, l'image fixe, la vidéo, [et] la communication interactive caractéristique du téléphone, s'intègrent [quant à eux] au sein du système technologique constitué par Internet. » De Rosnay désigne ainsi Internet comme un « système technologique complexe ».

L'ordinateur ne doit plus être considéré comme une machine à traiter de l'information, mais comme un outil de communication à part entière. Il est caractéristique de notre société de l'info. Peut-on imaginer aujourd'hui s'intégrer à la société, sans avoir accès à un outil multimédia ? Serait-on capable d'entrer en interaction avec notre environnement de manière aussi efficace ? J'en doute fort.

Pour exister au sein de la société de l'information, il semble indispensable d'être capable de partager ses propres ressources avec les autres dans le cadre du travail collaboratif, et aujourd'hui ce partage s'effectue naturellement via les TIC¹⁷.

Toujours dans cette idée d'une toute puissance de l'Internet sur nos échanges au quotidien et de sa capacité d'intégration des autres technologies, on observe un renforcement de la connectivité entre les différents outils qu'ils soient d'environnement différents (logiciels) ou de natures différentes (téléphones/smartphones, PDA, appareils photos, ordinateurs portables, Blackberry, caméscopes...).

Nous l'avons dit, une information a la propriété de se démultiplier. Ajoutons une nouvelle donnée : Internet offre la possibilité de rediriger les flux d'information. Je

¹⁶ La « WayBackMachine » est un moteur permettant de rechercher les pages Web ayant été indexées une fois mais qui depuis n'existent plus.

¹⁷ Cf. partie sur l'équipement des ménages.

peux par exemple faire une page Web sur laquelle je donne mon avis sur des événements d'actualité, une sorte de tribune. Ma tribune fera réagir d'autres personnes qui pourront répondre en ligne, et un flux d'information sera ainsi matérialisé. Si maintenant, je place sur ma page Web un hyperlien pointant par exemple sur un site raciste, on peut considérer que le contenu initial de ma page Web est maintenant surchargé par une nouvelle information, comme le principe de compositionnalité d'un mot au milieu de son contexte. Le mot « bombe » désigne quelque chose de différent dans la phrase : « la bombe qui a explosé en plein centre ville... », et dans la phrase : « cette fille est une bombe atomique... ». Le mot « étranger » sera sans aucun doute connoté de manière différente si la page Web qui le contient pointe vers le site du Front National.

Ce qui me permet de dire que le contenu d'une page sur Internet est par nature instable, et non figé.

Pour reprendre cet exemple, il faut savoir que si je peux décider de mettre un lien de ma page vers un autre site (lien interne), l'inverse est également vrai.

Ainsi, l'information qui se trouve sur mon site peut être connoté parce qu'un site raciste aura décidé de faire pointer un lien de son site au mien (lien externe).

Comment alors assurer au visiteur de ma page que mon propos initial n'est en rien xénophobe ?

Cet exemple montre à quel point nous nous devons d'être responsable du contenu que nous mettons en ligne. Il est difficile d'exercer un contrôle total sur les liens externes, et la plus grande vigilance doit être apportée aux liens internes que nous décidons de créer.

L'information a besoin d'être décodée, et il ne s'agit pas uniquement ici d'un problème technique. Plus d'information implique plus de médiateurs, de professionnels, journalistes et professeurs... pour l'analyser. Il faut rejeter l'information qui n'est pas validée.

Nous découvrons ainsi que si l'Internet est un espace qui permet de s'exprimer librement, il suppose aussi une prise de conscience de la part de ses utilisateurs, qui doivent être des contributeurs responsables.

II.4 La diversité du contenu présent sur le Web

II.4.a La construction du Web et de son industrie

C'est l'une des raisons du succès rencontré par le Web : la variété des informations qui s'y trouvent.

Si nous jouissons aujourd'hui d'une relative facilité pour accéder à ces informations et pour naviguer de liens en liens, cela n'a pas toujours été le cas, et comme toute révolution, la révolution Internet s'est construite pas à pas.

C'est grâce au commerce que nous avons pu accéder au vivier d'informations du grand réseau. Plus précisément, c'est le coup de génie qu'ont eu les FAI de faire de l'accès à l'information un commerce, qui a mis la machine en branle.

Les fournisseurs d'accès ont été les premières entreprises à réaliser du chiffre d'affaires sur Internet¹⁸. Sont ensuite apparus des logiciels dédiés permettant de tirer profit des liens hypertextes et de « surfer » sur cet étendue d'informations : ce sont les « navigateurs » comme Netscape ou Internet Explorer. Enfin sont apparus des logiciels permettant de centraliser les différentes routes d'information, en un seul point (c'est la naissance de l'idée de « portail »). Ces logiciels offraient la possibilité de rechercher des informations et surtout de les retrouver rapidement. Ainsi débuta le succès des moteurs de recherche et autres annuaires.

C'est cette idée de centralisation, de point d'entrée sur l'Internet préalable à toute orientation de l'internaute qui a fait des moteurs de recherche des acteurs incontournables du net.

Aujourd'hui nous nous orientons plus sur des services ajoutés grâce aux métamoteurs qui sont en quelque sorte, des portails de portails. N'oublions pas d'ajouter à cela une personnalisation grandissante des outils de recherche, de navigation et de redirection des flux d'information (agents intelligents, listes de diffusion, newsgroups...).

Dorénavant les utilisateurs d'Internet ne sont plus aveugles, ils peuvent voir ce que les moteurs de recherche retrouvent.

Il devient donc intéressant de se demander comment exister sur ce nouvel espace virtuel. Et si les fournisseurs d'accès ont été les premiers à comprendre le potentiel économique du cyberspace, cette dimension s'est rapidement imposée au point de devenir indissociable de la notion d'Internet.

Internet et le commerce entretiennent une relation fusionnelle. Et l'on imagine aisément les difficultés que soulève la coexistence dans le même espace, d'une gigantesque masse d'informations, et d'acteurs de plus en plus nombreux, motivés par la promesse d'un chiffre d'affaires alléchant.

II.4.b La promesse commerciale du Web.

L'utilisation de moyens électroniques pour des transactions commerciales et des échanges d'informations est un mouvement qui a été amorcé dans les années 60, essentiellement par les banques. Ce mouvement s'est développé dès les années 70 avec les standards EDI (Electronic Data Interchange), autorisant la transmission de commandes, de factures et d'ordres de livraison sur des réseaux de communication souvent privés.

Le développement des NTIC et de la nouvelle économie ont provoqué l'apparition de nouvelles activités économiques et ont modifié de nombreuses activités qui existaient déjà.

Dans le cadre de notre étude nous retiendrons la définition suivante du commerce électronique, qui est la définition retenue par l'OCDE:

« La vente ou l'achat de biens ou de services, effectués par une entreprise, un particulier, une administration ou toute autre entité publique ou privée, et réalisée au moyen d'un réseau électronique (computer-mediated networks) ».

¹⁸ Source : Joël de Rosnay

Cette définition inclut, non seulement les achats par minitel, par un système téléphonique interactif (type Audiotel¹⁹) ou encore, entre les entreprises, au travers de relations directes et automatisées d'ordinateur à ordinateur (type EDI²⁰).

Elle exclut les transactions passées sur un mode non interactif, par exemple par fax et téléphone, mais aussi les mails quand ils ne sont pas échangés dans le cadre d'une application interactive (message simple suivi d'un traitement manuel).

Par ailleurs, les biens et les services sont commandés par l'intermédiaire de ces réseaux électroniques mais le règlement financier et la livraison peuvent être réalisés en ligne ou par d'autres moyens.

II.4.c Un marché...potentiel

Avec la promesse de toucher une clientèle potentielle planétaire, la dimension informative du Web s'est donc rapidement doublée d'une dimension commerciale et il est devenu impensable pour tout commerce et toute industrie au sens large de ne pas y être représentés.

En mai 1998, on estimait à 115 millions, le nombre d'internautes dans le monde,²¹ entre 3,7 et 4,3 millions pour la France²² (ils seront environ 12 millions en 2001). Fin 2003, ce nombre est passé à 700 millions pour les internautes à travers le monde et à 24 millions pour la France.

Le nombre d'internautes au niveau mondial en 2004 a été estimé à 934 millions²³ ; et toujours d'après la même source, le milliard d'internautes devrait être dépassé en 2005 pour atteindre les 1,35 milliards en 2007.

II.4.d L'équipement des ménages

Selon l'Insee, en début d'année 2004, 31% des ménages français étaient connectés à l'Internet.

Cela s'explique d'une part, par la progression constante et régulière de l'équipement informatique des ménages français (+ 5 points de pourcentage chaque année)²⁴, et d'autre part, par la conversion rapide des internautes à l'Internet rapide (grâce à la baisse des tarifs d'abonnement « haut débit » (5 millions de connexions « haut débit » en France en juin 2004).

¹⁹ "composez le 3280 et dites: météo"

²⁰ Échange de données informatisées (Electronic Data Interchange). Transmission d'ordinateur à ordinateur, d'application à application, de données structurées selon des messages préétablis et normalisés via un moyen de télécommunication. Cette technique permet l'échange automatisé de données - codifiées et agencées selon un langage préalablement convenu - entre des applications logées sur des systèmes d'information distincts et hétérogènes. Les échanges sont effectués au moyen de différents réseaux de télécommunications.

²¹ Source : NUA, <http://www.nua.com/surveys/>

²² Source : Baromètre Multimédia de Médiamétrie, CSA, TMO, Médiangles - 2004

²³ Source : Computer Industry Almanac

²⁴ Source : SESSI

Nous avons évoqué plus haut la logique d'intégration selon laquelle fonctionne l'Internet. Toujours dans cette même logique, il faut savoir qu'à terme, tous les terminaux de communication deviendront des supports pour le commerce électronique, puisqu'ils seront tous interconnectés (71% des Français ont un téléphone portable et 22% ont accès à une offre de télévision numérique), de plus la téléphonie de troisième génération commence déjà à répondre concrètement à cette projection.

II.4.e L'équipement des entreprises

La diffusion des TIC dans les entreprises progresse : plus de 97% des entreprises françaises sont équipées en micro-ordinateurs, près de 83% d'entre elles sont connectées à l'Internet, dont 53% en haut débit (ADSL), et 83% utilisent le courrier électronique²⁵. Outre cet équipement de base en voie de banalisation, plus de 37% des entreprises ont mis en place un progiciel de gestion intégrée (ERP)²⁶ et 38% d'entre elles disposent d'un intranet destiné à optimiser la communication et donc le partage d'information en interne. De plus, grâce à des réseaux internes de communication comme les intranets, la répartition des pouvoirs s'effectue sur de nouvelles bases. Avec l'usage des intranets, les structures s'aplatissent, la circulation des informations se fait plus rapide. Grâce à la veille concurrentielle ou à l'intelligence économique, l'entreprise se dote d'un système sensoriel lui permettant de prévoir les évolutions et les crises. Sa réactivité aux modifications de l'environnement, grâce à la responsabilisation de chacun par l'information, lui confère un grand avantage compétitif. Le sentiment d'appartenir à une équipe et de voir émerger l'intelligence collective du groupe motive les employés et renforce l'esprit d'entreprise (travail collaboratif).

Pour optimiser les échanges avec leurs partenaires externes (fournisseurs, sous-traitants), environ 14% des entreprises ont mis en place un réseau extranet²⁷, et pour améliorer leurs relations commerciales avec leurs clients, 39% des entreprises sont présentes sur le Web au travers d'un site « en propre ».

Selon Forrester et Gartner, deux sociétés d'études spécialisées, le commerce en ligne à destination des entreprises représenterait des montants dix à quinze fois supérieurs au commerce à destination des particuliers.

Notons également que la diffusion des TIC dans les entreprises s'effectue avec un niveau d'équipement plus élevé dans les grandes entreprises, et d'autre part, que celles qui réalisent des innovations de produit sur le marché adoptent les TIC de manière plus significative.

²⁵ Selon l'enquête communautaire sur les TIC réalisée en fin d'année 2002

²⁶ Entreprise Ressource Planning, ou progiciel de gestion intégré. Logiciel permettant de contrôler les flux à l'intérieur des entreprises depuis les fournisseurs jusqu'aux clients (gestion commerciale, comptabilité, gestion de production, des ressources humaines).

²⁷ Réseau privé et confidentiel reliant deux ou plusieurs entreprises, et utilisant les protocoles et les outils d'Internet. Un extranet fait souvent communiquer entre eux des intranets. C'est un réseau externe utilisant la technologie IP (Internet Protocol). Il permet à une entreprise ou à un organisme d'échanger des informations numériques avec ses principaux correspondants (filiales, clients, fournisseurs, etc.) en bénéficiant de la norme IP pour la transmission des informations et d'une présentation conviviale des informations.

Le contexte de concurrence technologique dans lequel s'inscrit la production d'un site Web d'entreprise, contribue parfois à une désinformation de l'internaute : il est essentiel de convaincre les clients (finaux ou partenaires commerciaux) potentiels, que les produits et/ou services de l'entreprise sont les meilleurs ou deviendront les « standards » de leur industrie, ou d'un marché donné. La tentation est alors grande d'exagérer les informations affichées sur le site et qui concernent leurs parts de marchés réelles ou souhaitées²⁸.

II.5 L'autre impact des sites commerciaux

Faire du commerce ne se réduit pas au seul acte de vente et l'impact de l'Internet sur le commerce de détail ne se limite pas uniquement en terme de chiffre d'affaires généré en ligne. Cette précision permet aussi de mettre en exergue la pauvreté de l'expression « commerce électronique » qui en français ne parvient pas à rendre compte de l'ensemble des pratiques des entreprises, qui reposent sur des moyens et des plateformes électroniques (de la communication à la distribution en passant par la négociation) et qui sont regroupées en anglais sous la dénomination « e-business ».

L'e-business passe par la création de sites Internet qui diffusent des informations sur les produits et services de l'entreprise, passe par des intranets et des extranets.

On remarque en effet que les internautes utilisent les sites marchands comme des sources d'information. Ainsi, la navigation sur le site d'une entreprise, peut représenter une valeur ajoutée en terme d'apport de clientèle supplémentaire dirigée vers les points de vente physiques ou en terme de contacts commerciaux (conseils, information technique, pratique, etc.).

Le Web cesse d'être une place de marché globale (quelque peu fantasmée) pour devenir un media additionnel incontournable, permettant de diffuser de la publicité, de prospecter de nouveaux clients, de fournir de manière plus rentable des services additionnels, de gérer de façon plus dynamique des catalogues, des tarifs, la communication de l'entreprise.

Cette communication se retrouve au sein des quatre grands domaines de l'e-business que sont les activités destinées aux entreprises (business-to-business²⁹), les activités destinées aux consommateurs individuels (business-to-consumer³⁰), la communication de consommateur à consommateur (consumers to consumers³¹) ou la communication de consommateurs aux entreprises (consumers to businesses).

Le nombre d'internautes qui achètent en ligne est globalement comparable à celle des internautes qui préparent leurs achats en ligne et qui concluent leurs achats sur les points de vente physiques. A ce propos, il est amusant de noter l'émergence de certains particularismes nationaux: ainsi, les Allemands s'informent et achètent beaucoup en ligne, tandis que les Britanniques achètent beaucoup en ligne, mais se renseignent peu en prévision d'un achat hors ligne. La nature des produits achetés et

²⁸ Eric Brousseau « Commerce électronique : ce que disent les chiffres et ce qu'il faudrait savoir », 2001.

²⁹ ou "b-to-b", "B2B".

³⁰ ou "b-to-c", "B2C".

³¹ Les échanges entre consommateurs peuvent consister en des transactions via des sites de troc ou d'enchères (type Ebay) en plus des simples échanges d'informations (listes, newsgroup, email, etc.)

vendus peut expliquer une part de ces écarts: certains produits (voitures, maisons, etc.) sont encore difficiles à vendre sans contact physique avec l'objet. C'est pourquoi le Web semble mieux réussir au commerce de biens intangibles.

II.5.a Acheter sur le Web

Acheter en ligne devient de plus en plus un phénomène de société³².

En juin 2003, 36% des internautes âgés de plus de 18 ans et plus déclarent avoir effectué un achat en ligne au cours des six derniers mois³³, ce qui représente 7,5 millions de cyberacheteurs. Au niveau de leur profil, ces cyberacheteurs sont de gros consommateurs de l'Internet: 69% utilisent Internet quotidiennement; les non-acheteurs sont 43% à utiliser Internet de façon journalière.

Les différentes enquêtes précitées convergent vers une même conclusion: trois facteurs principaux conduisent à acheter en ligne:

- l'ancienneté de l'usage de l'Internet
- la fréquence de connexion
- la connexion à l'Internet à haut débit

et nous pouvons ajouter une quatrième raison qui n'est pas citée par ces études: le fait que l'acheteur dispose dans son réseau d'autres internautes ayant déjà fait des achats sur le Net... et s'en trouvant satisfaits.

Selon une étude universitaire menée aux États-Unis³⁴, le temps nécessaire d'utilisation de l'Internet avant d'acheter en ligne est de 36 mois et plus pour un tiers des internautes, entre 18 et 36 mois pour un cinquième d'entre eux, et entre 12 et 18 mois pour 15%.

Les personnes et les organisations ont besoin de temps pour s'habituer à ces nouvelles pratiques et il est bien connu que le consommateur a besoin de temps pour changer ses habitudes.

II.5.b Les principaux freins à l'achat sur Internet

La confiance dans l'achat en ligne, quant à elle se développe. En février 2004, une enquête montre que la méfiance des internautes à l'égard du paiement en ligne baisse lentement, à mesure qu'ils se familiarisent justement avec les achats en ligne³⁵. Mais en 2004 toujours, 51% des internautes (contre 55% en avril 2001), demeurent très méfiants et déclarent "ne pas avoir confiance et n'être pas disposés à acheter sur Internet". D'autant que les consommateurs eux-mêmes font circuler des informations, et sur le Web, la bouche-à-oreille (appelé « buzz ») va très vite, surtout lorsqu'il est négatif.

En Europe, un peu plus de 90 millions d'européens auraient acheté un bien ou un service sur Internet en 2004, selon Forrester Research. Le commerce électronique

³² Source :Tableau de bord du commerce électronique

³³ Source :baromètre Taylor Nelson Sofres, (cité dans Tableau de bord du commerce électronique)

³⁴ Source :UCLA, 2002 (cité dans Tableau de bord du commerce électronique)

³⁵ Source :enquête de CSA Opinion (citée dans Tableau de bord du commerce électronique)

représenterait en Europe 40 milliards d'euros, soit une augmentation de 47% par rapport à 2003.

En terme de vente en ligne aux particuliers (qui représentait 3,4 milliards d'euros de chiffre d'affaires en 2003 pour la France), les ventes les plus significatives concernent les voyages (44% en 2002), les achats informatiques (14%), l'alimentation (11,5%), les produits culturels (8,5), l'habillement (6%), et la billetterie (2%)³⁶.

Mais le commerce électronique concerne également des objets de plus grande importance comme des micro-ordinateurs ou encore des automobiles.

Toutes les activités commerciales sont concernées par la mise en réseau des activités économiques. Les grandes administrations utilisent également le Web pour communiquer avec leurs administrés, des États utilisent le réseau pour se faire connaître d'un point de vue touristique, industriel ou commercial.

Les hommes et les partis politique utilisent également ce media pour nouer des liens avec un électorat potentiel (on se rappellera André Santini, qui, lors du lancement de son site Internet <http://www.andre-santini.net/> était extrêmement fier de dire sur tous les plateaux télé et dans tous les magazines, que l'on pouvait le trouver sur le Web!).

³⁶ Benchmark Group 2002

Naissance des outils de recherche et fonctionnement.

Cette partie présente les moteurs de recherche en décrivant un à un chaque composant afin de démystifier leur modus operandi. Cette définition permettra également de préciser ce que les moteurs ne sont pas (moteurs, annuaires, etc.)

Nous observerons ensuite une photographie des technologies actuelles des deux moteurs mondiaux les plus populaires, les incontournables Yahoo ! et Google, en essayant d'expliquer leur spécificité.

III. Naissance des outils de recherche et fonctionnement.

III.1 Des problématiques croisées

Posons le petit problème de logique suivant :

A votre droite : des centaines de milliards de documents qui regroupent en vrac :

- les photos de vacances des Dubois au camping des quatre feuilles
- de l'information scientifique et technique
- des forums de discussion, des blogs, des emails sur la nouvelle pâte à tartiner X
- le site de la société qui fabrique la pâte à tartiner X sur lequel il est possible d'acheter la pâte à tartiner X et de s'informer sur la future pâte à tartiner X2.

A votre gauche : des milliards d'internautes (dans le sens : utilisateurs de la ressources Internet) et parmi eux :

- ✓ les Dubois, qui souhaitent que tout le monde puisse admirer leur bronzage rouge écrevisse
- ✓ des chercheurs qui souhaitent rendre accessibles leurs publications
- ✓ des personnes qui souhaitent donner leur avis sur un problème de société
- ✓ des commerçants qui veulent vendre, communiquer, acheter, se renseigner sur leurs concurrents
- ✓ vos fournisseurs (si vous êtes une entreprise) ou encore, vos concurrents.

Et puis vous ! Vous et votre besoin d'information, que ce soit un besoin de recherche ou de diffusion.

En l'absence d'outils dédiés, vous disposez de deux options pour retrouver l'information qui vous intéresse, ou pour permettre à l'internaute de retrouver l'information qui pourrait l'intéresser:

- 1) Soit la localisation du site sur le Web est connue, c'est-à-dire son adresse Internet, et on parle alors d'accès **direct**.
- 2) Soit il faut cliquer sur un lien situé sur un autre site, dans un email, ou dans une publicité (bannière, fenêtre pop-up, etc.), on parle alors de **lien d'accès ou d'accès externe**.

Alors comment faciliter l'accès à une information donnée quand l'internaute ne connaît pas l'adresse du site et qu'il ne peut décemment pas explorer l'hyperespace en espérant rencontrer un lien pointant vers le « bon » site ?...

C'est en cherchant une réponse à cette question que s'est naturellement imposée l'idée d'un outil de recherche, d'un guide qui serait capable de fournir à quiconque le consulte, les documents et/ou adresses Internet répondant à son besoin d'information. Ainsi a-t-on vu se développer, au sein des organisations et des entreprises des systèmes de gestion de l'information permettant l'indexation et l'interrogation automatique de leur fonds documentaire.

Ce succès, rencontré par les outils de recherche a toujours été lié à cette volonté de faciliter la vie de l'internaute (simplification de l'accès à l'information, de l'utilisation de l'outil, etc.). Et tout le développement des outils est allé en ce sens, du moins au niveau de l'interface de communication avec l'utilisateur final, car les développements et les algorithmes travaillant en boîte noire (non visibles par l'utilisateur), sont eux beaucoup plus complexes.

Nous utilisons à l'heure actuelle différents types d'outils de recherche mais les chiffres le prouvent : le moteur de recherche est plébiscité³⁷. Essayons de comprendre ensemble ce qu'est un moteur de recherche (et ce qu'il n'est pas) avant de démystifier l'objet en étudiant brièvement son mode de fonctionnement.

III.2 Définitions des notions de base.

III.2.a Qu'est-ce qu'un moteur de recherche ?

Un moteur de recherche est un programme informatique. L'objet du moteur de recherche est de retrouver à l'aide d'algorithmes³⁸ définis par le concepteur du programme, des documents, des informations, des URL³⁹ se trouvant dans une base de données, et qui répondent à une requête posée.

Par abus de langage, on associe souvent au terme de « moteur de recherche » le site Web dont la fonction première est de fournir un accès à ce programme informatique, dans le but de réunir et de renvoyer de l'information disponible sur Internet, ou plutôt sur une partie de l'Internet (cf. Taille du Web).

Ces sites Web utilisent un moteur développé par leurs soins (e.g. le site de la compagnie Google) ou acheté à un tiers (e.g. celui de Yahoo !). Les moteurs de recherche peuvent être complètement différents selon les documents qu'ils indexent sur le Web, et aussi selon leur manière de parcourir les index à partir de la requête de l'utilisateur.

Les moteurs utilisent des programmes qui visitent les pages Web et leurs liens de manière automatique afin d'indexer leur contenu.

³⁷On sait que "90,2% des internautes privilégient la recherche par mot clé à la navigation par rubrique thématique" Source : Livre Blanc Netbooster

³⁸³⁸ L'algorithme détaille via une suite d'instructions écrites dans un langage approprié, les opérations permettant de résoudre un problème.

³⁹ Acronyme de Uniform Resource Locator. Adresse Internet exploitée par les navigateurs (Internet Explorer, par exemple). En pratique, c'est l'adresse d'un site.

Sur le Web, la mise à jour régulière des index, permet aux moteurs de fournir rapidement des informations très récentes.

La rapidité du moteur à répondre à l'utilisateur dépendra pour sa part, des algorithmes mis en œuvre pour traiter documents et requêtes.

D'autres types de moteurs de recherche, utilisés en entreprise, effectuent leurs recherches sur les intranets⁴⁰, et depuis peu deux nouvelles sortes de moteurs sont apparus : les Toolbars (11 décembre 2000 Google Toolbar) et les Deskbars (HotBot Deskbar, Google Deskbar (for Enterprise), AskJeevesDeskbar, etc.

III.2.b Qu'est-ce qu'un portail ?

Un portail Internet est un point d'entrée sur le Web, à caractère généraliste, technique ou thématique. Il sert de guide aux utilisateurs pour trouver l'information ou un service d'une organisation particulière, utiliser les services électroniques et envoyer des questions et commentaires. Les portails fournissent un grand éventail de ressources et de services, y compris un répertoire d'autres sites Web, une facilité de chercher d'autres sites, des nouvelles, des courriels, de l'information téléphonique et géographique et quelques fois un forum de consultation publique.

Généralement utilisé comme terme marketing pour décrire un site Web qui est destiné à être le premier endroit que les gens voient lorsqu'ils utilisent le Web, un « portail » dispose, typiquement, d'un catalogue de sites Web, d'un moteur de recherche ou des deux. Un site portail peut également proposer la messagerie électronique et d'autres services pour encourager les gens à utiliser ce site comme leur principal "point d'entrée" (d'où le mot "portail") vers le Web.

Les portails «professionnels» rassemblent des informations plus ciblées, sous forme de liens, en adéquation avec la communauté visée. C'est une logique communautaire d'échange.

Ce nom est parfois donné aux moteurs de recherche, aux annuaires, à la page par défaut d'un navigateur, à la page de base du site d'un fournisseur d'accès, aux sites réunissant une sélection de liens triés par catégories, vers d'autres sites.

III.2.c Qu'est-ce qu'un métamoteur ?

Un métamoteur ou un métachercheur est un logiciel qui puise ses informations à travers plusieurs moteurs de recherche.

Un métamoteur est un moteur de recherche qui envoie la requête de l'utilisateur à plusieurs moteurs de recherche différents et qui renvoie les résultats de chacun de ces moteurs. Il permet à un utilisateur de saisir sa requête une fois pour toutes et

⁴⁰ Site Web interne à l'entreprise qui l'a développé, qui est réservé aux salariés autorisés, et qui utilise la même technologie que les sites Internet. Bien que les pages de l'intranet puissent s'ouvrir vers l'Internet, un intranet n'est pas un site auquel le grand public peut accéder.

d'accéder simultanément à plusieurs moteurs de recherche. L'idée est de gagner du temps et du rappel en conduisant plusieurs recherches simultanément, étant donné qu'il est difficile de cataloguer l'ensemble des contenus se trouvant sur le Web. La facilité d'utilisation et la forte probabilité de trouver l'information recherchée rendent les métamoteurs populaires auprès de ceux qui ont suffisamment de courage pour se frayer un chemin à travers les pages de résultats non pertinents.

Les métamoteurs ne compilent pas de bases de données. Ils n'indexent pas le Web. Ils fonctionnent comme des intermédiaires, entre la requête de l'utilisateur et les moteurs de recherche.

Chaque métamoteur a sa spécificité. Certains ne redirigent la requête que sur les moteurs de recherche les plus populaires tandis que d'autres incluent des moteurs moins connus, des newsgroups⁴¹ et d'autres bases de données. Une autre différence réside dans leur manière de présenter les résultats. Certains listent les résultats en fonction de l'algorithme de classement du moteur vers lequel a été redirigée la requête. Ce qui permet à l'utilisateur d'apprécier quel moteur de recherche répond à ses requêtes avec le plus de pertinence. Si l'utilisateur souhaite continuer plus avant, le métamoteur peut alors le rediriger vers ce moteur particulier et relancer la recherche sur la page d'accueil de ce dernier. D'autres classent les résultats en se basant sur les calculs de pertinence de l'ensemble des moteurs de recherche interrogés. L'avantage pour l'utilisateur est de disposer dans ce cas d'une liste de résultats à l'intérieur de laquelle les doublons ont été éliminés. Certains permettent même à l'utilisateur de choisir ses moteurs et de comparer les résultats « perdus » d'un moteur à l'autre, ce qui permet de se rendre compte de la couverture Web de chacun des moteurs.

Les moteurs de recherche attendent souvent les requêtes utilisateurs sous une syntaxe particulière. Par exemple, certains moteurs autorisent l'utilisation des opérateurs « ET » tandis que d'autres attendront « + » pour effectuer la même opération, d'autres encore n'auront besoin que d'un espace entre les mots, toujours pour effectuer cette opération d'union. Les meilleurs métamoteurs synthétisent les requêtes de l'utilisateur lorsqu'ils les soumettent aux moteurs de recherche.

Les résultats varient beaucoup d'un métamoteur à un autre. Quoiqu'il en soit, même le plus simple métamoteur permet d'explorer une plus grande partie du Web que n'importe quel moteur de recherche pris séparément⁴².

III.2.d Qu'est-ce qu'un annuaire...

Un annuaire est un outil de recherche qui recense des sites Web et les classe par catégories (on parle également de rubriques). Son fonctionnement peut être comparé à un annuaire téléphonique professionnel. Pour y figurer, un site Web est

⁴¹ En français groupe de discussion ou forums. Les forums sont des sites virtuels où les internautes peuvent déposer des courriers par thème; se rencontrer aussi pour partager leurs centres d'intérêt. Il existe des quantités de forums couvrant pratiquement tous les sujets possibles et imaginables. La plupart des fournisseurs d'accès à Internet propose une adresse de news qui est du type: news.nom du fournisseur d'accès.fr.

⁴² Ex. ariane6 metacrawler, kartoo, vivisimo, ez2find, FindForward, Ixquick, Copernic Meta, KillerInfo, SurfWax...

soumis à l'inscription dans la / les catégories souhaitée(s) au travers de fiches descriptives comprenant en général le titre, l'adresse URL et un bref descriptif allant de 15 à 25 mots maximum. Les annuaires sont gérés manuellement. En effet, les sites soumis sont validés (ou non) par des rédacteurs. Un utilisateur qui effectue une recherche peut taper une requête ou descendre directement dans l'arborescence des catégories.

(Ex ; Yahoo ! GoogleDirectory...)

III.2.e Qu'est-ce que le Text-Mining ?

C'est l'ensemble des techniques permettant l'extraction d'information à valeur ajoutée issue de grands corpus de données textuelles.

Le Text-Mining, aussi connu sous le nom d'analyse textuelle intelligente, de fouille de données textuelles ou d'extraction d'information à partir du texte (knowledge-discovery in text (KDT)), fait généralement référence à l'extraction d'information et de connaissances pertinentes et structurées à partir de données textuelles non structurées. Le Text-Mining est une discipline récente à la croisée de plusieurs domaines d'étude qui regroupe la recherche d'information, la fouille de données, les statistiques, la linguistique-informatique et le traitement automatique du langage. Étant donné que la majeure partie de l'information disponible sur le Web est textuelle, le text-mining semble promis à un avenir commercial certain.

III.3 Comment fonctionnent les moteurs de recherche ?

Les premiers moteurs de recherche sont apparus en 1995⁴³. De simples catalogues Web, ils sont devenus depuis des systèmes intelligents.

Le panorama des outils de recherche d'information se répartit aujourd'hui dans un triptyque : moteurs de recherche, annuaires thématiques et métamoteurs⁴⁴.

Le but d'un moteur de recherche, nous l'avons vu, est de fournir des documents pertinents répondant à une demande d'information formulée par un utilisateur.

Sans trop entrer dans la technique, essayons tout de même de comprendre comment fonctionne un moteur de recherche, et pour cela, décomposons un moteur de base (c'est-à-dire utilisant la simple recherche booléenne⁴⁵).

III.3.a Premier composant : L'interface graphique

La première étape de la recherche gérée par l'interface graphique, consiste à récupérer la demande formulée par l'utilisateur.

⁴³ Les premiers ne sont pas apparus sur le Web, cf. Annexe Historique

⁴⁴ <http://savoirscdi.cndp.fr/CulturePro/actualisation/Serres/Serres.htm>

⁴⁵ Permet de combiner entre eux des critères de recherche pour sélectionner directement les références que l'on cherche et elles seules. Ces combinaisons s'expriment à l'aide de trois opérateurs booléens : ET, OU, SAUF (AND, OR, NOT) cf. chapitre dédié.

Pour communiquer avec le programme informatique de manière conviviale, il est préférable que l'utilisateur puisse agir via une interface homme-machine (IHM), qui se traduit aujourd'hui presque toujours sous la forme d'une interface graphique. Cette interface graphique doit contenir au moins un champ permettant à l'utilisateur de saisir du texte, et un bouton d'action, lui permettant de lancer une recherche à partir du texte saisi.

Certaines interfaces graphiques autorisent l'utilisateur à effectuer une « recherche avancée », dans ce cas plusieurs champs de saisie et/ou menus déroulants et/ou cases à cocher, permettent de préciser les critères de recherche. (Selon les études d'usage de l'Internet et des moteurs effectuées par Médiamétrie eStat et @position, seul 1.08% des utilisateurs français s'aventurent dans les onglets « recherche avancée » du moteur de recherche « Google » (Google est le moteur le plus utilisé par les français avec 72.69% d'utilisateurs en France et 56.4 % au niveau mondial en Mai 2004⁴⁶). En Amérique du Nord, ils sont 21% tous moteurs confondus, à affirmer utiliser la recherche avancée de manière fréquente.

La seconde étape prise en charge par l'interface graphique, est bien entendu l'affichage des résultats. Le moteur doit m'informer du succès (ou de l'échec) de ma requête.

En cas d'échec, l'interface peut m'indiquer que ma requête n'a pas aboutie (c'est la moindre des choses). Elle peut également me proposer d'effectuer une nouvelle recherche, me faire des suggestions pour mieux formuler ma demande. En cas de succès, différents modes de visualisation et de consultation des documents existent, quant à l'accès au document, le plus simple est de pointer vers lui à l'aide d'un lien hypertexte.

III.3. b Deuxième composant : L'index et l'opération d'indexation

L'index est un fichier ou un ensemble de fichiers qui regroupe des informations sur les documents contenus dans la base, on le nomme aussi « fichier inverse »

A l'image de l'index que l'on peut trouver à la fin d'un livre (les anglophones parlent de « back-of-book-style indexing »), il regroupe des couples [clé|valeur] permettant d'associer à une donnée certains attributs (typiquement, un mot associé à un numéro de page (par opposition à une recherche taxinomique (par hiérarchie de catégories))).

Les index permettent d'accélérer la vitesse de la recherche à l'intérieur d'un document ou d'une base documentaire.

Lors d'une recherche sur un moteur Web, le programme informatique ne parcourt pas l'ensemble des pages Web à la recherche des mots contenus dans ma requête (cela prendrait énormément de temps) mais il parcourt ce fichier, ou cet ensemble de fichiers, ce qui explique la relative rapidité d'exécution des recherches.

⁴⁶ Source : www.indicateur.com

L'index est le fruit d'une opération : l'indexation, qui, lorsqu'elle est manuelle, est régie par une norme AFNOR de 1978. Il n'existe cependant aucune norme pour les moteurs de recherche, ce qui illustre l'anarchie pure des nouvelles technologies. Selon Pomart et Sutter, l'indexation a pour but de faciliter l'accès au contenu d'un document ou d'un ensemble de documents.

- **En quoi consiste l'indexation ?**

La définition suivante s'applique plus particulièrement à l'indexation manuelle, car nous admettrons qu'à ce stade, les moteurs de recherche ne sont pas capables de catégoriser l'information.

L'indexation manuelle, donc, consiste à analyser les documents afin d'en extraire les notions qui s'y trouvent. Puis, ces notions doivent être transcrites dans un langage normalisé, afin de pouvoir être :

- Facilement accessibles par le plus grand nombre d'utilisateurs.
- Facilement exploitables par un système informatique.

Cette transcription des notions dans un langage normalisé permet de rendre compte du contenu du document. Mais les données permettant de localiser le document sont tout aussi indispensables et doivent donc obéir aux mêmes règles.

En effet à quoi servirait un moteur de recherche ou un documentaliste (dans le cas d'une indexation manuelle) qui vous affirme qu'il a bien un document qui répond à votre demande mais qui ne sait ni quelle forme il a, ni où il se trouve ?

- **Indexations**

Il existe deux types d'indexation : d'une part, l'indexation manuelle et d'autre part, l'indexation automatique.

Chaque type d'indexation a ses avantages et ses inconvénients.

L'indexation manuelle demande énormément de temps, d'argent, et des compétences étendues de la part des indexeurs humains qui en ont la charge.

Le principal avantage de l'indexation manuelle lorsqu'elle est contrôlée, c'est-à-dire lorsqu'elle s'appuie sur un thésaurus (sur une liste de termes faisant figure d'autorité), est de permettre à l'indexeur humain de retrouver sans difficulté le document qu'il a lui-même indexé.

Mais cet avantage est rapidement noyé sous la masse de difficultés que soulève ce type d'indexation :

Premièrement, si la personne qui a indexé un document le retrouvera sans difficulté majeure, toute autre personne aura beaucoup de mal à retrouver le document, les usagers étant tous limités par les décisions du documentaliste.

Donner un accès au thésaurus à l'utilisateur est donc une première manière de l'aider dans sa recherche.

Deuxièmement, le temps et l'argent requis pour indexer manuellement un document sont estimés [Chaumier, 90] respectivement à 6mn et 7.5€(50francs) par document. Le coût moyen pour une base de données regroupant 100.000 documents est donc de 750 000€ et de 416 jours dédiés à l'indexation, plus les frais générés par les documents devant être mis à jour chaque année.

L'indexation manuelle est donc une tâche lourde étant donné la masse de documents, mais c'est aussi une tâche qui requiert de nombreuses compétences de la part de l'indexeur : en effet, il lui faut

- comprendre le ou les domaine(s) abordés dans les documents,
- maîtriser les différentes langues dans lesquelles ces documents sont rédigés,
- il lui faut d'autre part maîtriser les langages documentaires
- ainsi que le processus de transfert de l'information.

En pratique, le documentaliste reçoit un document :

Il lui attribue :

- un numéro d'archivage afin de pouvoir le localiser ultérieurement
- une notice bibliographique afin de rendre compte du support, c'est ce que l'on nomme le catalogage.
- Et enfin a lieu l'indexation proprement dite (la transcription du contenu du document).

Pour cette indexation du contenu, les documentalistes procèdent à une « lecture diagonale », c'est-à-dire qu'ils parcourent le document, en s'appuyant principalement sur sa table des matières. Il s'en servent pour extraire les notions principales qui, selon eux, sont contenues dans le document, puis transcrivent ces notions ou concepts⁴⁷ en utilisant les termes (on parle alors de « descripteurs ») se trouvant dans leur liste de mots de référence : le thésaurus.

Le choix des descripteurs dépend:

- Du type d'information recherché
- De la finesse demandée pour les requêtes (granularité : requêtes unitermes, formulaires détaillés)
- De la disponibilité et de la compétence des indexeurs
- Et du domaine (généraliste ou spécialisé).

L'indexation automatique, elle, vise à effectuer les opérations d'archivage, de catalogage et d'indexation de contenu à l'aide d'outils informatiques, afin de produire un index sans intervention humaine (sinon pour coder le programme ce qui n'est pas négligeable).

L'indexation automatique peut elle-même être décomposée en deux catégories :

D'une part l'indexation automatique non sélective et d'autre part l'indexation automatique sélective.

⁴⁷ (Lat. conceptus, de concipere " recevoir "). Représentation abstraite d'un objet ou d'un ensemble d'objets ayant des caractères communs. Ex : concept « oiseau » (vole, possède un bec, etc.)

- **L'indexation automatique non sélective :**

L'indexation automatique non sélective ou plein texte (de l'anglais « fulltext ») permet de générer un fichier inverse à partir de tous les mots contenus dans le document indexé.

Et ce avec toute l'ambiguïté que soulève la notion de « mot ».

Par exemple : combien de mots la phrase « *La pomme de terre que j'ai mangée au restaurant d'entreprise aujourd'hui m'a donné des haut-le-cœur.* » compte-t-elle? Certains diront 21, d'autres 14, d'autres 12, etc.

Une fois réglée la question de ce qu'est un mot et de ce qui en constitue les frontières, les documents sont découpés en mots ou « tokens »⁴⁸ puis à chaque mot est associé un ensemble de valeurs, définies dans les spécifications du programme d'indexation (n° de document, de page, de ligne, position dans la phrase, etc.).

On associe généralement au fichier inverse ainsi constitué un anti-lexique (appelé stop-list en anglais) qui permet de ne pas indexer les mots considérés comme étant peu discriminant dans le cadre d'une recherche d'information (habituellement les mots vides (mots grammaticaux du type déterminant, conjonction, etc.)). Cette façon de procéder était celle de Google durant une longue période mais il semblerait que le moteur prenne de nouveau en compte les mots grammaticaux dans les requêtes.

L'indexation automatique non sélective (fulltext) implique de disposer d'un corpus de travail numérique.

L'intérêt de cette indexation plein texte est de permettre d'accéder de façon plus précise à ce que l'on recherche. Ainsi, même si un document ne parle pas en général d'un certain thème, il peut contenir une information précise sur ce thème à laquelle le fulltext me permet d'accéder. Par exemple, un document parlant des insecticides ne sera pas classé dans coccinelles mais pourra préciser que les coccinelles sont utilisées comme insecticides naturels, ce qui pourra intéresser un chercheur effectuant une recherche sur les coccinelles.

Les circuits documentaires sont optimisés grâce à la rapidité de traitement liée à cette méthode. Un autre avantage non négligeable tient au fait que le vocabulaire utilisé pour décrire le contenu du document est issu du document lui-même, avec pour conséquence une uniformité et une régularité de l'indexation sur des documents identiques que ne permet pas l'indexation manuelle (dans le cas par exemple, ou deux indexeurs humains auraient à décrire le même document à l'aide du même thésaurus)

Mais si cette indexation plein texte basique présente des avantages évidents par rapport à une indexation manuelle, force est de constater qu'elle génère de nombreux effets inattendus qui en font un procédé économique mais insuffisant.

Les insatisfactions liées à ce type d'indexation découlent de son principe de base: les textes sont découpés en mots, mais dans le cadre actuel du traitement, tel que nous venons de le présenter, un mot n'est qu'une suite de caractères, au sens informatique

⁴⁸ Cette opération répond au nom de « tokenization ». En programmation, un token est un segment de texte, qu'il soit lisible ou constitué de symboles. C'est la plus petite unité lexicale.

du terme, c'est-à-dire une suite d'octets, une sorte de code barre, dénué de sémantique⁴⁹ et de règles morphosyntaxique⁵⁰

Ainsi la recherche d'information est à ce stade plus proche d'une recherche de motif (une sorte de « grep » sous Unix): je ne cherche pas réellement le **mot** « table », mais un T. suivi d'un A suivi d'un B , etc. Pour être plus précis, nous pouvons même dire que, pour le logiciel, la suite T.A.B.L.E n'est qu'un code binaire, c'est-à-dire une suite unique de 0 et de 1... Le logiciel est donc loin de savoir que cette suite peut aussi bien désigner un objet concret (table à repasser) qu'un objet abstrait (table de multiplication) ou la forme conjuguée d'un verbe (ex : tableur sur).

Pour dresser un parallèle avec le reproche qui a été fait précédemment à l'indexation manuelle, l'utilisateur n'est plus limité par les décisions du documentaliste, mais, dans le cadre d'une indexation plein texte sans autre fonctionnalité, il se voit limité par la langue même du document, et cette fois, il n'a pas la possibilité de se référer à un thésaurus qui offre des solutions de parcours thématiques et d'utilisation de synonymes.

Les difficultés que rencontre alors l'utilisateur sont les suivantes :

- Le bruit, autrement dit, la quantité de documents ne répondant pas de manière pertinente à sa requête, est important, puisque rien ne permet à l'index d'obtenir des indices sur le sens des mots.
- Le silence, c'est-à-dire la quantité de documents pertinents qui se trouve dans la base et que le système n'a pas renvoyés à l'utilisateur est également important, puisqu'une simple marque de pluriel (un s à la fin de « table ») constitue dans sa logique, un mot (ie. un code binaire) différent.
- Toujours en ce qui concerne le silence, les documents pertinents qui ne reprennent pas « mot pour mot » les termes saisis dans la requête utilisateur, mais qui sont sémantiquement liés à ma requête (synonymes, formulations différentes, etc.) ne seront pas non plus renvoyés.

D'autre part, si le programme renvoie bien les documents contenant les suites exactes de caractères soumises dans ma requête, il ne possède aucune instruction lui permettant de m'indiquer quel document renvoyé est susceptible de mieux correspondre à cette requête, ni aucun moyen de m'éviter la lecture de documents identiques (logique booléenne).

L'idée est donc d'améliorer, de compléter le processus d'indexation automatique plein texte afin de le rendre sélectif : c'est-à-dire, d'essayer de rendre le système capable de « juger » de la pertinence des résultats qu'il trouve dans sa base documentaire.

D'une part, en comparant les résultats entre eux pour les présenter à l'utilisateur du plus pertinent au moins pertinent, et d'autre part, en appliquant à la base documentaire une série de traitements qui permettront au système de récupérer le plus grand nombre d'indices possibles et d'enrichir ces chaînes de caractères qu'il liste dans son index.

Ces améliorations contribueront à l'amélioration du système et à un appariement plus pertinent entre requête et documents.

⁴⁹ D'une manière générale, la sémantique est une branche de la linguistique qui étudie les signifiés, autrement dit le sens des mots. Le mot sémantique a été inventé à la fin du XIX siècle par le linguiste français Michel Bréal, auteur du premier traité de sémantique.

⁵⁰ En grammaire, analyser un segment de discours consiste à évaluer, d'une part la forme (morphologie flexionnelle), d'autre part la fonction (syntaxe) de ses éléments constitutifs.

- **Le « Web indexing »**

L'indexation Web ou indexation Internet regroupe l'indexation des sites Web ou des sites intranet via un fichier inverse, et la création de métadonnées (important, les métadonnées, à définir) servant de mots-clés et permettant de fournir un vocabulaire plus puissant, plus utile à des moteurs de recherche Internet, ou fonctionnant en local sur le site Web.

L'indexation via un fichier inverse diffère d'un autre type d'indexation que l'on rencontre également sur certains sites, et qui est la taxinomie, cette fois, la navigation à travers l'index n'a pas lieu via l'ordre alphabétique, mais en parcourant différentes strates de catégories hiérarchisées.

- **Pourquoi indexer manuellement un site ?**

L'indexation manuelle d'un site est souvent confondue avec ce que l'on appelle le référencement.

Le référencement est l'ensemble des interventions visant le positionnement stratégique optimal d'un site Web sur les services de recherche, pour les mots-clés les plus pertinents. Le référencement est une discipline complexe, fluide et dynamique, qui requiert la présence d'un spécialiste, et ce, dès la phase de la conception du site.

Le référencement est lié à un travail d'optimisation.

Il permet de choisir les mots clefs adaptés à chaque moteur.

L'indexation automatique ne peut tenir compte de la concurrence. L'analyse préalable à l'indexation manuelle permet de placer le site efficacement parmi les sites concurrents. Le choix des mots clefs et des codes META se fera en tenant compte des autres sites et de leur efficacité.

L'indexation manuelle ne se contente pas des seuls moteurs généralistes à forte concurrence. Elle utilise les index thématiques, propres à promouvoir efficacement le site. Car qualifier le trafic, c'est accroître l'efficacité du site.

Il vaut mieux parler d'indexation manuelle pour l'inscription manuelle des sites par les opérateurs humains, notamment au sein des annuaires thématiques comme Yahoo ! Dans ce cas, c'est l'opérateur qui affectera les mots clés qui lui auront semblé être les plus pertinents pour qualifier le site.

III.3.c Troisième composant : Le module d'appariement

Le module d'appariement est la partie du moteur de recherche, (autrement dit l'algorithme du logiciel de recherche) qui est dédiée à la récupération des termes saisis par l'utilisateur, à leur analyse, et à leur confrontation avec les termes contenus dans l'index, c'est la fonction de recherche du moteur.

Le rôle du module d'appariement est de comprendre et de traduire en temps réel **les mots et les idées** de l'utilisateur vers l'index (et par extension, la base documentaire) et si possible du document vers l'utilisateur.

La langue utilisée pour communiquer avec un moteur de recherche, est beaucoup plus simpliste que celle que nous utilisons tous les jours pour communiquer. Les requêtes sont en général courtes (« une recherche type est composée de 2 à 3 mots et plus de 50% des recherches sont composées de 3 mots et plus »).

Le challenge à relever pour ce module est donc de permettre à l'utilisateur de trouver ce qu'il cherche. Ce qui implique de collecter tous les indices possibles aussi bien à l'intérieur de la requête qu'à l'intérieur des documents (+ ressources lexicales externes... à voir plus tard dans le cadre des améliorations de l'indexation et de la classification automatique).

La mise au point d'un module d'appariement efficace passe donc par l'observation et le décryptage attentifs des schémas de communication d'humain à humain en vue de les adapter au schéma de communication d'humain à machine.

En résumé : le locuteur conceptualise son message, puis l'exprime sous une forme orale et/ou écrite et/ou visuelle. L'interlocuteur reçoit le message, le décrypte à partir des indices qui s'y trouvent, qui se trouvent dans le contexte de la communication, et dans sa connaissance personnelle du monde. Ces indices lui permettent de rattacher le message du locuteur à des notions (concepts). Le schéma est donc :

Conceptualisation → Cryptage → Décryptage → Conceptualisation → Cryptage, etc..

La machine, elle, ne pense pas, elle ne possède pas de connaissance du monde qui lui permette de rattacher les mots à des concepts. Elle possède uniquement des mots, qu'elle tente de faire concorder.

La communication homme – machine implique donc :

1. soit d'utiliser le « langage » de la machine : c'est-à-dire de ne pas partir de nos concepts et de notre connaissance du monde, mais plutôt du vocabulaire de la base documentaire, seule « connaissance » du moteur de recherche
2. soit de permettre à la machine de rattacher nos mots (qui sont chargés de notre connaissance du monde) à une connaissance du monde, à une représentation de concepts qui seront les siennes.

III.3.d Quatrième composant : La base documentaire

Un corpus est un ensemble de documents (textes, images, vidéos, etc.), regroupés dans une optique précise. On peut utiliser des corpus dans plusieurs domaines : études littéraires, linguistiques, scientifiques.

Ses caractéristiques doivent être prises en compte pour optimiser son exploitation par un moteur de recherche (quelle taille, quelle fréquence de mise à jour, quel format des documents, type de documents, type d'informations contenues, langue (s) des documents, localisation (plusieurs bases, une seule machine, etc.)).

III.3.e Cinquième composant : L'utilisateur

L'utilisateur est la donnée la plus importante de l'équation. Simplement parce que c'est autour de sa satisfaction que doit se construire un moteur.

Le moteur doit donc pouvoir gérer :

- les langues de l'utilisateur
- les aptitudes de l'utilisateur, son degré de familiarité avec l'informatique, avec les moteurs de recherche, avec la spécialité du corpus...
- sa localisation géographique (où est l'utilisateur ? sur un ou plusieurs sites ?)
- la fréquence d'interrogation de l'outil, la précision attendue, les fonctionnalités souhaitées, etc.
- son type d'interface favori (deskbar, toolbar, page Web...)

III.4 La composition des moteurs, en résumé :

Un moteur de recherche compte cinq paramètres de base sur lesquelles il est possible d'agir en vue d'améliorer l'efficacité de l'outil :

- l'interface graphique (à l'intérieur de laquelle se distinguent une interface d'interrogation d'une part, et une interface de présentation des résultats d'autre part)
- l'index (taille, format.)
- et son processus de création : quelle forme est indexée (mot lui-même, stem, lemme...), quels attributs (informations) sont associée(s) aux formes : document, position dans le document, forme d'origine, temps d'indexation, etc.
- le corpus (la base documentaire) : quelle taille, quelle fréquence de mise à jour, quel format des documents, types des documents, type d'informations contenues, langues des documents, localisation (plusieurs bases, une seule machine, etc.)
- le traitement de la requête par le module d'appariement (ou « fonction de recherche ») : temps de recherche, informations utilisées (comme les connaissances linguistiques)
- l'utilisateur (pédagogie, etc.)

III.5 État de l'art de la technique en matière de recherche d'information

A l'heure actuelle la recherche d'information se limite à une recherche de motif, à comparer une chaîne de caractères avec une autre chaîne de caractères.

On s'oriente vers un Web sémantique, l'identification précise des ressources grâce aux URI, la structuration logique des ressources grâce à XML, une description complète, structurée et pertinente des ressources grâce aux métadonnées RDF, à une représentation partagée des domaines de connaissances : les ontologies avec OWL.

III.5.a Approche générale

Les moteurs Web stockent des informations concernant un très grand nombre de pages Web, puisées directement à partir du WWW. Ces pages sont recensées par un Webcrawler (littéralement : celui qui rampe sur la Toile), aussi surnommé spider (araignée).

Il s'agit d'un navigateur Web automatisé, configuré pour suivre tous les liens hypertextes qu'il rencontre sur son chemin. Le contenu de chaque page est alors analysé afin de déterminer de quelle manière cette dernière doit être indexée (par ex. : des mots sont extraits du titre de la page, des noms de chapitres, ou de champs spécifiques : les balises méta).

Les données des pages Web sont stockées dans une base de données (l'index), elles serviront à répondre à de futures requêtes.

Certains moteurs, tel que Google, conservent une image complète ou partielle du code source de la page (nommée « cache » dans les pages de résultats de Google) ainsi que des informations sur les pages, tandis que d'autres stockent tous les mots de toutes les pages rencontrées (comme par exemple le moteur Altavista).

L'utilité de la page stockée en cache est qu'elle permet de consulter un résultat contenu dans une page, même si cette page a été enlevée du Web ou que son contenu a été modifié sur le site d'origine : une manière de combattre la dégénérescence des liens (linkrot = liens morts ou obsolètes).

Comme pour les moteurs de recherche qui ne sont pas dédiés à la recherche sur Internet, un utilisateur entre sa requête en soumettant des mots-clés. Le moteur parcourt ensuite son index et fournit une liste des pages Web les plus pertinentes, selon ses propres critères en proposant généralement une fiche composée du titre du document et parfois de quelques lignes de texte (snippet).

L'utilité d'un moteur de recherche dépend de la pertinence des résultats qu'il renvoie. Bien qu'il puisse y avoir des millions de pages Web contenant un mot ou une phrase en particulier, certaines pages peuvent être plus pertinentes, plus populaires, ou plus dignes de confiance que d'autres, mais comment mesurer ces notions de pertinence, de popularité et de confiance ?

La plupart des moteurs de recherche Web utilisent des méthodes spécifiques pour classer leurs résultats afin de fournir à l'utilisateur les « meilleurs » résultats en tête de liste, que l'on appelle « le haut du panier ».

Ces méthodes varient largement d'un moteur à un autre, mais également à travers le temps, parallèlement aux usages de l'Internet et à l'évolution de nouvelles techniques.

Petit panorama des algorithmes des deux plus importants outils de recherche mondiaux.

III.5.b Google

Présentation

Développé par l'université de Stanford par Sergey Brin et Larry Page, deux jeunes de 26 ans diplômés en science informatique, Google a fait son apparition on-line, fin septembre 1999, après plus d'une année de tests.

Google est le moteur de recherche le plus populaire, leader mondial des outils de recherche (avec 56.4% de parts de marché en Mai 2004⁵¹).

Le succès de Google tient au fait qu'il trouve toujours quelque chose et qu'il le trouve vite. Son interface est simple et dépouillée, ce qui est rassurant pour le néophyte.

Google supporte plus de 104 langues ou dialectes et propose une version personnalisée de son moteur pour 113 pays (en comptant l'interface www.google.com), la recherche pouvant être restreinte au Web de 38 pays.

Deux types d'accès sont ainsi possibles selon l'interface :

1. • Par langue nationale, régionale ou dialecte (en vert)
2. • Par pays ou fédérations (en bleu)

Google est connu pour son moteur de recherche mais il propose également un annuaire (Google Directory) basé sur l'annuaire Open Directory (projet d'annuaire collaboratif) appelé DMOZ.

Le succès de Google s'est construit sur une communication originale, qui le démarque de ses concurrents : loin des médias traditionnels (excepté les objets publicitaires disponibles dans la boutique Google (« GoogleStore »)), Google base sa stratégie de communication sur les communiqués de presse, le buzz (bouche à oreille), et la communauté des développeurs.

L'interface est simple et colorée avec des couleurs qui rappellent la pâte à modeler, et de petites subtilités telles que les logos personnalisés à différentes périodes de l'année, représentent autant de petites attentions qui touchent l'affectif de l'utilisateur en l'accompagnant au fil des jours et qui le fidélisent... sans pour autant amener quoique ce soit à l'efficacité pure du service.

⁵¹ Source OneStat « baromètre mondial des outils de recherche » (Juin 2002 à Mai 2004)

- **La pertinence selon Google**

Google est « populaire » car il utilise de façon judicieuse la notion de « popularité » sur le Web pour classer ses résultats. Pour Google, une page Web pointe sur une autre page Web lorsqu'elle lui trouve un intérêt. « Ce préjugé est typique de la culture universitaire des créateurs de Google. Dans le milieu académique, on cite toujours par déférence, pour souligner l'apport et la qualité de ceux qui nous ont précédés. »

L'indice de popularité de Google, est selon ses concepteurs, un modèle du comportement de l'utilisateur. C'est indice porte le nom de « Page Rank » :

« Nous nous donnons un surfeur virtuel à qui l'on fournirait une page au hasard. Il clique de lien en lien sans jamais revenir en arrière (bouton 'back') mais, quand il s'ennuie, il demande une nouvelle page de départ (toujours tirée au hasard). L'indice PageRank d'une page donnée représente la probabilité que notre surfeur la visite. La pondération (d) est la probabilité que cette page ennue le surfeur et qu'il en demande une autre⁵². »

Le PageRank : « Supposons la suite de pages T1...Tn pointant vers une même page A. Le paramètre d est un indice de pondération compris entre 0 et 1 (nous le réglons généralement sur 0,85). La valeur C (A) représente le nombre de liens sortant de A.

L'indice PageRank de la page A se calcule comme suit :

$$PR (A) = (1-d) + d (PR (T1)/C (T1) + \dots + PR (Tn)/C (Tn))$$

A ce calcul s'ajoute l'**analyse de mise en correspondance hypertextuelle** : Google associe le texte du lien à la page vers laquelle il pointe.

- **La limite du modèle.**

Le phénomène de popularité est récursif : le plébiscite est d'autant plus fort qu'il émane d'une page elle-même bien cotée avec pour conséquence un renforcement des positions dominantes. Les nouveaux sites n'y trouvent donc pas d'intérêt.

- **La petite bête de Google :**

Le moteur d'indexation de Google s'appelle « Googlebot » (« bot » pour « robot »). Ce Webcrawler parcourt les pages, les analyse, et suit chacun des liens qu'il y trouve. L'indexation des sites par Googlebot est gratuite, quelle que soit la nature du site indexé.

⁵² Source : Livre Blanc Inffluo « Moteurs de recherche, où est la technologie ? »

III.5.c Yahoo !

Né en 1994 de deux doctorants de Stanford David Filo et Jerry Yang dont le hobby était d'indexer les sites Web qu'ils jugeaient intéressants, le « Guide du World Wide Web de Jerry et David » est devenu **Yahoo !** [Yet Another Hierarchical Officious Oracle!] . Depuis, **Yahoo!** est devenu une marque internationale incontournable avec des bureaux à travers le monde (<http://docs.yahoo.com/info/misc/history.html>).

Puis **Yahoo!** a racheté **Inktomi** en 2002 et **Overture** en 2003, elle-même propriétaire de **AlltheWeb** et de **Altavista**.

Si Yahoo ! est toujours premier mondial au niveau du trafic des sites de son réseau (avec 121 millions de visiteurs uniques et un temps de connexion moyen par internaute de plus d'une heure et 20 minutes⁵³), il est après Google en terme de trafic généré par son outil de recherche⁵⁴.

Yahoo ! est un annuaire mais cette fonctionnalité s'efface de plus en plus pour faire place à un moteur de recherche. Le moteur de recherche que propose Yahoo ! sur sa page d'accueil a longtemps été basé sur le moteur de Google : Yahoo ! avait signé un partenariat avec Google pour exploiter sa base de données et ses indications de classement, auxquelles Yahoo ! ajoutait des références issues de son propre annuaire. Depuis mars 2004, En 2004, **Yahoo!** a abandonné la technologie du moteur Google et lancé son moteur de recherche basé sur la combinaison des techniques de recherche des ses récentes acquisitions (technologie baptisée: la Yahoo Search Technology (YST)) et fournissant un service accordant plus d'importance au moteur de recherche ainsi constitué qu'à l'annuaire qui avait fait sa réputation au commencement.

⁵³ source Nielsen/NetRatings « Le classement mondial » Mai 2003

⁵⁴ Google était classé 4^e pour le trafic des sites en mai 2003 avec 83 millions d'internautes et un temps moyen de 18 mn

Pourquoi les outils de recherche ont-ils tant d'intérêt pour le commerce électronique ?

Pourquoi le commerce et les moteurs de recherche se sont-ils rencontrés ?
Interrogeons-nous sur le véritable rôle de ces outils et sur la manière dont nous les utilisons. Que représentent-ils pour les sociétés commerciales ?
En quoi notre manière d'échanger de l'information a-t-elle changé ? Et quelles sont les tendances vers lesquelles les outils de recherche évoluent ?
Enfin, pouvons-nous appliquer la notion de gratuité concernant l'information disponible sur le Web ?

IV. Pourquoi les outils de recherche ont-ils tant d'intérêt pour le commerce électronique ?

Selon une étude californienne, 16 % du temps de travail en entreprise serait consacré à rechercher l'information. *"Soit, si l'on transpose à un travailleur français soumis aux 35 heures, un peu plus de 5 heures 30 par semaine..."*⁵⁵ Il est donc indéniable que la recherche d'information possède un coût économique et qu'elle représente bien un enjeu majeur. D'autre part, alors qu'on dénombre au niveau mondial sur Internet, plusieurs dizaines de millions de sites Web⁵⁶, être visible sur la toile représente plus que jamais un challenge.

Quel rôle jouent donc les outils de recherche à côté des autres formes de promotion et comment les internautes les utilisent-ils ?

IV.1 Le rôle des outils de recherche

Les outils de recherche jouent un rôle majeur dans la découverte de nouveaux sites. Face à l'explosion du volume d'information disponible sur le Net, la publicité sur médias classiques ou sur Internet connaît une très forte progression. De nombreux sites se sont en effet engagés dans cette voie. En parallèle, au sein du NetMarketing, le référencement se révèle un outil essentiel dont il convient de cerner les enjeux.

En effet, 99,2% des internautes français utilisent les outils de recherche. C'est ce qui ressort d'une étude en ligne qu'a menée Benchmark Group en juillet-août 2000 auprès de 1 481 internautes.

IV.1.a Comment les internautes utilisent-ils les outils de recherche ?

Il existe en théorie deux types d'outils de recherche. D'une part, les annuaires qui permettent une recherche par navigation dans des rubriques et sous-rubriques thématiques. D'autre part, les moteurs de recherche qui reposent sur un principe d'interrogation par mot clé. Dans la réalité, la majorité des outils de recherche combinent les deux approches (Google, Yahoo, Altavista, MSN...).

Or le mode d'utilisation des outils de recherche joue énormément sur la visibilité des sites Web.

En effet, dans les rubriques et sous-rubriques des annuaires, les sites sont classés par ordre alphabétique. Il est donc extrêmement difficile de garantir une meilleure visibilité de son site, si ce n'est en misant sur la qualité de son intitulé.

⁵⁵ Certaines sources parlent de "neuf heures et demie par semaine en moyenne, pour les cadres" (Management, N° 125, novembre 2005).

⁵⁶ Les estimations varient selon les études, retenons le chiffre comptabilisé par Netcraft à partir des noms de domaine : 75 millions de sites Web pour octobre 2005 (cité sur www.silicon.fr/getarticle.asp?ID=11953).

A l'opposé, les moteurs de recherche offrent une réelle marge de manœuvre puisque c'est le niveau d'adéquation entre le mot clé tapé et le contenu des sites qui conditionne l'ordre d'affichage sur les pages de réponses.

Autre donnée intéressante : 90,2% des internautes privilégient la recherche par mot clé à la navigation par rubrique thématique, ce qui confirme le plébiscite des moteurs de recherche.

Dans un contexte où la recherche par mot clé prédomine très largement, l'ordre d'apparition des sites dans les pages de réponses joue un rôle majeur. En effet, plus de 70% des internautes ne vont pas au-delà des 30 premières réponses que proposent les moteurs de recherche à la saisie d'un mot clé.

En conclusion, les moteurs de recherche représentent une source de trafic considérable pour les sites. Cependant, l'utilisation qu'en font les internautes est très pénalisante pour les sites qui ne figurent pas, dès le départ, dans les réponses proposées par les moteurs. C'est pour cette raison que le référencement revêt un caractère tout à fait stratégique, car c'est de sa qualité dont dépendra la visibilité d'un site.

L'enregistrement sur les outils de recherche est désormais devenu incontournable, et différentes solutions permettant l'automatisation de l'inscription d'une URL auprès de plusieurs centaines de moteurs de recherche dans le monde sont apparues sur le marché.

IV.1.b Attrait en terme de trafic généré et de communication

Internet s'est imposé comme une source d'informations privilégiée. Aujourd'hui, c'est aussi à la communication que cette fonction profite. Selon une étude présentée par l'Union des annonceurs (UDA), la société de services e-marketing Isobar et Overture, spécialiste du lien sponsorisé, 92% des internautes sont déjà allés sur le site d'une marque ou d'un produit à la suite d'un message publicitaire. 45% d'entre eux ont consulté dans la journée le site mentionné. Et les moteurs de recherche s'imposent comme «relais des médias traditionnels», dans la stratégie de communication digitale des entreprises.

«Il existe une influence forte entre les campagnes télévision ou radio et les recherches effectuées sur le réseau», indique Bertrand Jonquois, directeur exécutif d'Overture. Des pics de fréquentation ont ainsi été constatés sur Internet lors de la diffusion au même moment de campagnes hors ligne, notamment pour Dove («Pour toutes les beautés»), Transatlantys (voyages-sncf.com) ou encore Scenic (Renault). Selon l'étude UDA/Isobar/Overture, une majorité des internautes sont passés par un moteur de recherche après avoir vu ou entendu un message «hors ligne». Huit internautes sur dix consulteraient un site comme Google ou Yahoo ! après la mémorisation partielle des adresses de sites.

« Pour Pierre Calmard, codirecteur général d'Isobar France, «les internautes qui cherchent sont les prospects les plus intéressés». L'annonceur doit donc chercher à

«impliquer le consommateur et générer du désir» plutôt que d'«exposer une marque et générer de la notoriété».

La communication sur Internet exige de se pencher sur les stratégies de recherche des internautes. Les mots clés font plus souvent référence à une problématique («insecticide», «film action») qu'à la marque ou à l'entreprise. Quitte à inclure de tels termes dès la création publicitaire, en fonction de ce qui sera susceptible d'être recherché sur le Net. «Tant que les consommateurs iront chercher sur les moteurs de recherche, c'est là qu'il faudra aller», résume Bertrand Jonquois. Il est temps de prendre le train de la «googlisation» en marche »⁵⁷.

IV.1.c Attrait en terme d'accès à l'information

Les moteurs de recherche sont les vecteurs de l'information. Les liens renvoyés par Google qui est le moteur le plus utilisé finissent par paraître le renvoi d'une vérité validée. Ce que je veux dire c'est que Google est une fenêtre sur les vérités et contrevérités qui circulent sur Internet, et que le propre de ce moteur est d'ouvrir sa fenêtre sur une vue qui est la même pour tous, une vue qui répond à une doctrine académique selon laquelle les travaux les plus cités seraient les plus proches de la réalité.

Si l'algorithme de Google décidait demain d'ajouter à son algorithme un critère favorisant le champ lexical du « proaméricain », c'est un contrôle politique que le moteur exercerait alors sur l'information.

IV.2 Une nouvelle manière d'échanger de l'information

Jusqu'à présent, les deux modes classiques de relations s'appuient sur l'administration et la télévision. L'administration permet de faire remonter une demande personnelle vers une " machine administrative " à traiter cette information. Le moyen le plus efficace pour toucher un très grand nombre d'utilisateurs à partir d'un petit groupe de personnes est la télévision. Cependant, la remontée de l'information vers le centre d'émission ne se fait qu'avec un faible rendement. Internet permet aujourd'hui des communications de groupes à groupes. On peut atteindre rapidement un individu isolé et donc personnaliser l'information. Mais on peut également toucher une multitude d'individus par des listes d'envoi, des forums ou " *news groups* ". Autre modification profonde des usages : l'accessibilité à l'information. Le mode classique de consommation d'informations se réalise par la diffusion des émetteurs vers les récepteurs. C'est le cas, bien sûr, de la télévision. C'est aussi celui de la radio ou de l'édition, qu'il s'agisse de livres, de journaux ou de magazines. L'utilisateur adopte alors une attitude passive. En revanche, sur Internet l'utilisateur a la possibilité de créer de l'information. De passif il devient actif et même interactif. Il devient émetteur d'informations et peut explorer l'information proposée par les autres. Ces pratiques remplacent la logique traditionnelle de diffusion, par une logique de navigation et donc de responsabilisation face à la masse des informations proposées.

⁵⁷ Le Figaro du 11/07/2005

IV.2.a La spécialisation des moteurs

La tendance actuelle des outils de recherche d'information est à la spécialisation. On trouve désormais sur Internet des outils de recherche pour à peu près tous les supports, tous les thèmes, toutes les régions, toutes les applications possibles et la spécialisation est bel et bien une tendance « lourde » de cette industrie. Ce qui pose le problème de la recherche de ces outils spécialisés eux-mêmes, justifiant ces « méta-outils », ou répertoires d'outils, comme <http://www.enfin.com> (qui recense 6316 outils) ou Internet Search Engine Database.

La spécialisation revêt toutes les formes possibles :

- sur un domaine particulier (tourisme, industrie, culture, médecine, sciences exactes, sciences humaines et sociales, etc.)
- sur une zone linguistique ou géographique, selon la nature des documents (forums, listes de diffusion, bases de données, thèses, dépêches d'actualité, bibliothèques électroniques)
- sur le type de fichier (format)
- sur la nature du media (images, sons, etc.)
- des modes de visualisation des résultats différents (clustering⁵⁸, pictogrammes⁵⁹, bulles⁶⁰, carrés⁶¹)

Cette spécialisation touche toutes les familles d'outils : annuaires, moteurs, métamoteurs. Chaque catégorie se divisant entre outils généralistes et spécialisés.

- **Les comparateurs de prix**

Les comparateurs de prix vous permettent de trouver sur Internet et ainsi comparer, d'un simple clic, les prix de vente de produits pratiqués sur différents sites de commerce électronique.

Afin d'apprécier la qualité du site et la pertinence de la sélection des offres présentées, vous devez prendre soin de vérifier notamment les éléments suivants :

- le nombre de sites commerciaux pris en compte : les comparateurs de prix ne couvrent pas tous les sites de commerce électronique disponibles sur la toile mondiale. Ils se limitent généralement aux enseignes françaises les plus connues et plus rarement les sites aux audiences plus faibles.
- la réactualisation de la base de données : les offres présentes sur les sites de e-commerce évoluent en permanence, notamment au travers de promotions ou réductions opérées par ces commerçants. En l'absence d'une mise à jour rapide, vous pourriez ne pas prendre connaissance d'une proposition tarifaire plus intéressante.
- l'indépendance du site vis-à-vis de ses partenaires commerciaux : généralement gratuits pour le public, les comparateurs de prix peuvent se rémunérer notamment par la conclusion de partenariats avec certains sites ou

⁵⁸ www.exalead.com/search

⁵⁹ www.kartoo.com

⁶⁰ www.grokker.com

⁶¹ www.marumushi.com/apps/newsmap

au travers de l'affichage de banderoles publicitaires. Pensez à vérifier si ce partenariat n'est pas susceptible d'affecter les offres présentées.

De son côté, l'Union fédérale des consommateurs (UFC - Que Choisir) a publié une étude invitant les consommateurs à la plus grande prudence dans l'utilisation des comparateurs de prix ; l'association relevant une certaine opacité sur leur fonctionnement⁶².

IV.2.b La personnalisation des moteurs et les réseaux sociaux

Après sa solution d'historisation des recherches ("My Search History"), proposée il y a quelques semaines de cela, Google a lancé une solution de personnalisation baptisée "Personalized Search"⁶³. En fait, cette fonctionnalité existait depuis un certain temps dans les "Google Labs". La nouveauté est que les deux outils sont maintenant corrélés : l'outil de personnalisation "apprend" un peu plus de vous au fur et à mesure de vos recherches sur le Web. Par exemple, une requête sur le mot clé "moteur" fournira des liens vers Google, Yahoo! et MSN (ou autres) si vous avez effectués par le passé de nombreuses recherches au sujet des moteurs de recherche. Les résultats seront plutôt orientés vers Renault, Ferrari et Peugeot si votre passé dénote une tendance plutôt portée vers la mécanique automobile... La recherche personnalisée deviendra donc de plus en plus pertinente au fil de vos requêtes sur le moteur...

Yahoo!, de son côté, n'est pas en reste avec l'outil "MyWeb" en version 2.0⁶⁴, qui mélange réseaux sociaux et recherche d'information. Chaque utilisateur de l'outil d'historisation MyWeb, peut ajouter des mots clés et des descriptifs aux pages qu'il sauvegarde puis partager ses informations avec les autres utilisateurs. Pour l'occasion, Yahoo! aurait développé un nouvel algorithme de pertinence nommé "MyRank", plus spécialisé dans la recherche à l'intérieur de corpus de données communautaires. Une notion de partage d'information et de "réseau social", déjà ébauchée sur des outils comme Eurekster⁶⁵, certainement essentielle dans les moteurs de recherche de demain...

⁶² Source : http://www.droitdunet.fr/par_profils/lecture.phtml?type=profil_consommateur&it=5&id=4

⁶³ <http://www.google.com/psearch>

⁶⁴ <http://myWeb2.search.yahoo.com/>

⁶⁵ <http://www.eurekster.com/>

IV.3 Le Web et la culture du gratuit

De plus en plus se répand l'idée que l'information, en raison de son abondance, doit tendre vers la gratuité (les infos à la radio ou à la télévision sont déjà gratuites pour ainsi dire, seuls les journaux papier et les articles de certains quotidiens en ligne sont encore payants)⁶⁶.

Le Web a renforcé cette idée en faisant de la gratuité une culture : n'importe qui peut désormais y trouver l'information qu'il cherche, et chacun peut diffuser l'information qu'il veut. Le lecteur participe ainsi à l'écriture de l'article qu'il lit et peut construire lui-même la matière de ses informations de manière interactive, en sautant librement d'un document à l'autre, grâce aux liens hypertextes.

Le journaliste est conduit à renvoyer plus ou moins systématiquement aux différents sites dans lesquels il a puisé l'information. Sa fonction est donc menacée puisque, sur le Web, chaque surfer peut endosser la casquette de journaliste.

Les journalistes ne sont plus les producteurs exclusifs de l'information et ne peuvent plus revendiquer vraiment leur qualité d'intermédiaires entre le public et les émetteurs de l'information. Chaque individu peut devenir, théoriquement, un média.

Pire, nous sommes entrés dans une logique de self-service appliquée à l'information, les professionnels se trouvant de plus en plus réduits à "merchandiser", l'internaute faisant ensuite son choix, malheureusement trop souvent, comme dans la grande consommation, en privilégiant l'emballage au contenu (beauté du site, etc.) puisque les consommateurs y sont sensibles⁶⁷.

C'est en cela que la recherche d'information a connu une véritable révolution: dorénavant, les journalistes ne remplissent plus cette fonction d'intermédiaires, de relais entre les émetteurs d'information, et les usagers.

Ces derniers sont passés d'une situation de dépendance totale vis-à-vis des professionnels à une interaction directe avec les outils.

- **Des Interférences entre commerce et moteurs Web**

De nombreuses pages Web sont mises à jour régulièrement ce qui oblige les moteurs à les revisiter fréquemment.

Les requêtes que l'on peut soumettre à un moteur fonctionnent à partir de mots-clés, ce qui peut avoir pour conséquence de générer du bruit dans les résultats renvoyés.

Les sites générant un contenu dynamique peuvent ralentir ou rendre difficile l'indexation ou même conduire à une surproduction de résultats extraits d'un même site.

Certains moteurs ne classent pas les résultats en fonction de leur pertinence, mais en fonction des contributions financières que les sites ont versées pour apparaître plus haut dans la liste de résultats.

⁶⁶Source : <http://ethique.ifrance.com/hypermarche.html>

⁶⁷ Comme l'affirment les internautes interrogés dans le cadre de l'étude Consumer WebWatch.

Modèles d'affaires des moteurs de recherche

Nous définirons ici la notion de Keyword Marketing après avoir présenté le contexte qui a permis son apparition.

Nous verrons de quelle manière ce modèle a été décliné à différentes applications marketing afin de financer les sociétés éditrices de moteurs de recherche.

Pourquoi est-il important pour les sociétés de figurer dans les bases de données des moteurs de recherche, et qui sont les principaux acteurs du Keyword Marketing ?

Nous verrons également les différentes méthodes de référencement « borderline » ou abusives. Comment les internautes et les Webmestres parviennent-ils à contourner les modèles, ou à les détourner dans le but de lancer des « attaques informationnelles » comme le « Google Bombing », ou la « Fraude au clic » ?

V. Modèles d'affaires des moteurs de recherche

Nous avons bien vu que la taille du Web est aujourd'hui encore une donnée très difficile à évaluer (Web visible et Web invisible). Une chose est sûre: pour exister sur le Web, pour y être visible, mieux vaut être référencé au sein des moteurs de recherche. Mais un simple référencement (« naturel ») ne suffit pas toujours; ce référencement doit être optimisé. Les sites souhaitant être vus et visités doivent viser le “haut du panier” des pages de résultats renvoyées par les moteurs de recherche lors de la saisie d'une requête, c'est-à-dire les 30 premiers résultats (2 ou 3 premières pages).

Avant de définir le Keyword Marketing, voyons quel contexte a rendu ce modèle économique possible et rentable pour les exploitants des SRI⁶⁸.

Voici quelques chiffres qui nous permettrons de mieux comprendre le contexte du Keyword Marketing:

Selon une étude Carat Expert⁶⁹, 92% des internautes sont déjà allés sur le site d'une marque ou d'un produit suite à un message publicitaire, principalement (58%) dans le but d'obtenir des informations complémentaires. Pour autant la mémorisation des adresse sur les medias traditionnels demeure partielle et 62% des internautes en moyenne déclarent avoir des difficultés à mémoriser les adresses Internet mentionnées dans les publicités et plus de 8 internautes sur 10 utilisent naturellement un moteur de recherche suite à une mémorisation partielle des adresses de sites⁷⁰. 40% des internautes n'utilisent généralement qu'un seul moteur de recherche. Les 2/3 des personnes sondées déclarent utiliser les moteurs de recherche pour leur efficacité (gain de temps), les raisons suivantes avancées étant:

- avoir des informations complémentaires
- éviter d'avoir à saisir l'adresse URL des sites
- visiter d'autres sites

Les stratégies de recherche portent davantage sur l'entreprise (40%) que sur le produit ou service faisant l'objet d'une publicité (22%).

Ne pas trouver le site de l'annonceur altère l'image de la marque dans plus d'un cas sur trois.

Cette étude conclut sur le postulat suivant: Contrôler sa présence dans les premiers résultats des moteurs de recherche est un complément incontournable des campagnes publicitaires sur les medias traditionnels.

⁶⁸ Systèmes de recherche d'information

⁶⁹ Source Carat Expert « Les moteurs de recherche, relais des médias traditionnels »

⁷⁰ Plus d'une femme internaute sur trois a tout le temps recours à un moteur de recherche ainsi que 40% des 25-34 ans.

Ce postulat se voit renforcé par l'information suivante, extraite d'une autre étude⁷¹ qui affirme que plus de 70% des internautes ne vont pas au-delà des 30 premières réponses que proposent les moteurs de recherche suite à la saisie d'un mot-clé. Sachant que les pages de réponses des moteurs comportent généralement entre 10 et 20 réponses chacune, un référencement optimal doit permettre au site d'apparaître sur la première ou sur la deuxième page de résultats.

Le caractère stratégique du référencement est démontré: "c'est de sa qualité dont dépendra la visibilité d'un site".

V.1 Le Keyword Marketing ou le commerce des mots-clés

Le caractère stratégique du référencement ayant été démontré et étant perçu à la fois par les entreprises et par les sociétés d'outils de recherche, de nouveaux modèles commerciaux se sont développés en relation directe avec le mode de fonctionnement, l'algorithmique des moteurs de recherche.

Jusqu'alors, les modes de communications des sociétés, et les modèles économiques des moteurs se limitaient respectivement en l'achat et en la vente d'espaces publicitaires, de bannières commerciales présentes sur les pages d'accueil des moteurs.

Aujourd'hui, liens sponsorisés, liens commerciaux, "AdWords"... les techniques de ce qu'on appelle le "*positionnement payant*" ou "*Keyword Marketing*" ne cessent de se diversifier et de se développer, ajoutant un nouveau défi et un nouveau risque pour les usagers parfois distraits : savoir distinguer du premier coup d'œil un lien "sponsorisé" d'un résultat "normal".

Le Keyword Marketing consiste en un système de vente aux enchères de mots-clés, par des sociétés spécialisées (comme Overture) ou certains moteurs de recherche (comme *Google*). Cette vente de mots-clés permettra par exemple à un site commercial, spécialisé dans le voyage, d'apparaître en haut d'une page de résultats pour toute requête comprenant le mot "voyage". Avec le positionnement payant, c'est la notion même de pertinence qui est atteinte.

V.1.a L'innovation du coût par clic.

En 1998, Overture (ex GoTo.com et racheté en juillet 2003 par Yahoo ! pour 1.63 milliards de dollars en actions et en comptant) a eu l'idée d'introduire un nouveau système de revenu aux moteurs de recherche : le coût par clic (pay per click).

Avec ce nouveau système, il devient possible de placer une enchère sur un mot, une expression, une phrase, etc. de son choix (en 1998, quelques centimes pour un mot). Plus le mot clé est tapé sur les moteurs, plus celui-ci est prisé par les annonceurs et donc plus les enchères sont élevées.

Ensuite, à chaque fois qu'un internaute saisira ce mot, cette expression, etc. dans la fenêtre du moteur de recherche, un lien vers mon site apparaîtra sur la page de

⁷¹ Livre Blanc Benchmark Group "Les enjeux du référencement de sites Internet" consulté en ligne le 08/10/2005 à 20h00 sur http://www.journaldunet.com/livres_blancs/netbooster/netbooster.pdf

résultats. A ce moment-là, si l'internaute clique sur mon lien (qui est alors un lien promotionnel, puisque son référencement a été acheté), je devrais payer le montant de l'enchère que j'avais placée sur le mot (ou moins que le prix prévu, le prix étant fonction de la compétition). En général, plus l'enchère est élevée, plus le positionnement du site est proche du « haut du panier » sur la page de résultats.

Ce système a rapidement évolué pour aboutir à l'un des modèles d'affaires les plus attractifs de l'Internet.

- **Overture**

Overture était le plus grand moteur de recherche coût par clic avant que Google ne lance son système baptisé « AdWords ».

Overture fonctionne sur un système strict d'enchères : plus on mise, et mieux l'on est positionné.

Les résultats achetés auprès d'Overture sont reportés sur les résultats sponsorisés de Yahoo!, MSN, Lycos, Hot Bot et d'autres moteurs encore. Overture affirme toucher 80% des internautes au niveau mondial.

- **Google AdWords**

Le programme de coût par clic made in Google, « AdWords », a débuté en Février 2002, et est rapidement devenu le concurrent numéro un du programme leader d'Overture.

Le système de Google AdWords est différent du système d'Overture en ceci que l'enchère n'est qu'une partie de l'équation du positionnement dans les classements.

L'autre partie de l'équation est le taux de clic (en anglais : click-through-rate), c'est-à-dire, le nombre de fois où un internaute clique sur le lien menant à votre site.

Le système de Google AdWords approvisionne les pages de résultats des moteurs AOL, Netscape, AskJeeves, Teoma, Earthlink, et Google bien sûr.

Google affirme que ses « AdWords » apparaissent 200 millions de fois par jour. Évidemment, ce sont les données des entreprises elles-mêmes et il y a donc fort à parier que Google « atteigne également environ 70% des internautes du monde ».

Pour souscrire au service AdWords de Google, les annonceurs doivent acheter un mot-clé à Google France (par exemple). Chaque mot a une valeur marchande qui varie selon la demande entre 5 centimes et plusieurs euros. « Hôtel » par exemple, valait 3 euros début septembre (c'est moins qu'au printemps, lorsque se préparent les vacances), alors que « crédit » le plus cher de tous, valait 6 euros⁷².

La place des annonces commerciales dépend de deux critères : le montant investi et le nombre des clics des internautes sur l'annonce.

⁷² Source Capital octobre 2005

- **Google AdSense**

AdSense⁷³, une technologie dérivée a ensuite étendu ce système d'annonces ciblées à tous les sites Web même ceux des particuliers (par ex. Alain Debuisson jardinage.com empoche plus de 1000 euros par mois grâce aux annonces renvoyant vers des fleuristes) Google « offre » son moteur aux sites Web et les rémunère même afin qu'il fassent le relais vers les partenaires commerciaux de Google. Si vous vous trouvez sur un site perso et que vous effectuez une recherche sur crocodile. Le moteur ajoutera aux résultats disons, un lien commercial vers un partenaire commercial de Google exerçant dans la maroquinerie. Si vous cliquez sur ce lien, Le maroquinier paie une contrepartie à Google, et Google, reverse une partie de ses gains à l'auteur du site.

Une tendance lourde, parfois insuffisamment connue, se développe depuis plusieurs années et est en passe de transformer en profondeur le paysage de la recherche d'information : l'inscription, au cœur même des procédures de recherche, de la commercialisation. Il ne s'agit pas de la fin d'un prétendu âge d'or de la gratuité de l'information, qui n'a jamais vraiment existé. L'information a toujours eu un coût (le plus souvent supporté par le fournisseur) et souvent un prix pour l'utilisateur. Les banques de données professionnelles et scientifiques sont le plus souvent payantes (et souvent très chères), et ce depuis leur naissance.

La nouveauté réside ici dans cette nouvelle forme d'économie et de marché, apparue autour des outils de recherche privés du Web et des enjeux financiers énormes, à la mesure du trafic généré par ces outils et que nous venons d'expliquer.

D'autres moteurs ont développé leur propre système de coût par clic, mais nous ne retiendrons que ces deux systèmes car ils détiennent l'essentiel du marché.

Déjà en 2002, 14% du chiffre d'affaires de Yahoo! provenait de son partenariat avec Overture, bénéficiant ainsi de la croissance du chiffre d'affaires global de la publicité on-line lié aux liens sponsorisés.

V.1.b La soumission payante ou « insertion payante »

Certaines sociétés paient pour que leur site soit examiné plus souvent que les autres. Cette technique du référencement payant, utilisée il y a encore peu de temps par Overture (Altavista, AlltheWeb...) " garantit une simple insertion dans la page de résultats, sans promotion particulière". Les dirigeants de Yahoo ont lancé une offre de référencement payant. S'il est facile d'imaginer qu'en utilisant le terme **voiture** on verra apparaître des liens promotionnels (Yahoo), des liens commerciaux (Google) ou des liens sponsorisés (Voila), nous pouvons être surpris des résultats obtenus avec **enseignant ou médecine préventive ou cancer**. "Même les requêtes les plus banales peuvent parfois susciter des résultats peu conformes à l'objet initialement recherché. Sur de nombreux moteurs de recherche, les requêtes lancées avec le mot clé "fille" ou "garçons" peuvent rapporter, au gré des indexations, de nombreux sites à caractère pornographique, dont les pages comportent ces termes".

⁷³ Le programme AdSense de Google a été lancé officiellement en Mars 2003

V.2 Les méthodes de référencement alternatives, abusives⁷⁴:

Nous l'avons vu, les résultats qui s'affichent dans les pages renvoyées par les moteurs lors d'une recherche subissent le classement d'algorithmes puissants, dans le meilleur des cas et parfois, ce classement est également fonctions de différentes interventions pécuniaires comme le positionnement payant.

Toutes les sociétés ne peuvent pas se permettre financièrement de surenchérir face aux grosses entreprises lors des enchères de mots-clés ni s'offrir des insertions payantes. De nouvelles techniques ont donc vu le jour, formes de solutions parallèles pour référencer son site au sein des moteurs de recherche, sans déboursés un centime : Comme le dit Marc Duval " Connaître les techniques de référencement, c'est savoir pourquoi une page est affichée parmi les 10 premières même si elle ne contient aucun mot d'une requête en apparence. C'est aussi une façon de jauger la qualité d'un automate de recherche."

De nombreux procédés sont utilisés :

V.2.a Le détournement de l'utilisation de la balise <Title>

D'ordinaire ce titre, est très informatif. On sait que les algorithmes d'indexation de la plupart des outils de recherche tiennent largement compte du contenu de cette balise. Ainsi la société <http://www.sofymetal.com> a indiqué dans cette balise : <Title>SOFYMETAL ACIERS, aciers fins, aciers spéciaux, acier disponible sur stock, stock d'acier personnalisé, flux d'acier d </Title], renforçant par la répétition du terme **acier(s)** ses chances d'apparaître dans les premiers résultats avec le mot clé acier au singulier comme au pluriel. Cependant, les moteurs de recherche semblent désormais considérer cette technique comme abusive (spamdexing).

V.2.b L'utilisation d'informations invisibles (« Le texte invisible »).

Il s'agit d'écrire des mots clés en tout petits caractères, ou au contraire de choisir de gros caractères et de les afficher avec la même couleur d'écriture que le fond de page. Ainsi un site sur la Martinique pourra comporter 40 fois le mot clé **Martinique** dissimulé en haut de page (il faut sélectionner toute la page pour voir démasquer le subterfuge, ou consulter le code source pour lire le contenu caché).

⁷⁴ Regroupées sous l'appellation « spamdexing » (indexation-pollution)

V.2.c Le cloaking (ou technique des pages masquées)

Il consiste à créer deux versions de la page. La première est destinée aux robots indexeurs des moteurs de recherche et est "optimisée" par des mots clés, des liens... La seconde est destinée à l'utilisateur final (qui ne voit, lui, que celle-ci).

V.2.d Les liens hypertextes dissimulés

En cachant des liens hypertextes avec la balise <noscript> (normalement destinée à afficher un message si la version du navigateur ne permet pas d'exécuter un petit programme inclus dans la page), il est possible de tromper l'algorithme de Google. Qui s'appuie largement sur la présence de liens hypertextes dans la page considérée et sur l'existence de liens extérieurs pointant vers une page (ou backlink : lien entrant).

V.2.e L'utilisation de redirections trompeuses

L'objectif est d'attirer vers un site n'ayant rien à voir avec la requête de l'internaute. Des requêtes du type service+ville (par exemple patinoire Colombes) conduisent à de multiples pages qui, toutes, renvoient au même site. Cette pratique est pourtant dénoncée par Google. Par ailleurs, il est à noter que la plupart de ces pages citent Google (dans leur code source) comme "le moteur le plus utilisé" et que le site vers lequel ces pages convergent dispose de liens sponsorisés sur le même Google.

V.2.f L'utilisation de liens au sein d'un réseau (le « Google Bombing »)

Il s'agit de "tromper" Google en s'appuyant sur son algorithme d'indexation. Comme on vient de le voir, ce moteur tient compte des liens pointant vers un site. Le PageRank (ou indice de popularité) de Google intègre ce critère dans le calcul de la pertinence d'une page. "véritable clé de voûte de Google, [il] fonctionne également sur le principe de la communauté. Plus il y a de gens compétents dans votre domaine qui parlent de vous et plus vous êtes « populaire » dans votre domaine". Cet indice tient compte de la popularité d'un site et est associé à d'autres critères moins importants. Si cette popularité peut être un gage de fiabilité de l'information du site, comme une bibliographie est un élément d'appréciation d'un ouvrage; elle n'est, pour certains rédacteurs, qu'une technique pour arriver en première position dans les résultats de Google. "Certains professionnels n'ont pas hésité à "exploiter cette faille en créant artificiellement des pages de liens sans pertinence ou en développant des stratégies « agressives » d'échange de liens". Comme le dit Olivier Andrieu, l'algorithme de Google est devenu suffisamment sophistiqué pour tenir compte, "non seulement de la quantité de liens pointant vers une page, mais également de l'indice de popularité des documents qui les affichent.", autrement dit pour s'appuyer sur la quantité (nombre de liens) et la qualité (sites eux-mêmes très populaires). Néanmoins, on peut constater que la pratique du Google bombing est à la mode. Ainsi, **miserable failure** renvoie (ou a renvoyé) sur le site de la Maison blanche, **magouilleur** renvoyait, fin mars

2004, vers le site de l'Élysée. Récemment, un concours de référencement s'est engagé visant à être en première position dans les résultats de Google avec l'expression **mangeur de cigogne**. Entre le 23 mars et le 26 avril, le nombre de réponses obtenues avec ce moteur est passé de 80 100 à 240 000. Si l'on considère la seule recherche **mangeur de cigogne**, on voit bien les possibilités de tromper un moteur de recherche, même si l'on peut en sourire. Cependant, il n'est pas possible de faire une recherche sur la cigogne sans aboutir sur les sites de ses prétendus prédateurs car le mot clé **cigogne** donne environ 10 fois plus de réponses que la normale.

- **Conclusion**

En conclusion, les stratégies de référencement pèsent lourd dans l'indexation effectuée par les robots tout comme dans l'affichage des résultats. Il est difficile, sans les connaître, de prendre de la distance avec les réponses trouvées. Distinguer les liens commerciaux, savoir que certains sites peuvent faire leur promotion en utilisant des méthodes plus ou moins honnêtes, peut permettre un regard plus critique sur les réponses des moteurs de recherche.

Outre la création de nouvelles pratiques d'attaques informationnelles les algorithmes des moteurs de recherche, les internautes mal intentionnés s'en prennent parfois directement aux hommes et aux organisations via le lancement de rumeurs (« hoax »), sur tous types de médias Internet (forums, listes de diffusion, emails, blogs, etc.) et sur des pratiques plus surnoises comme la fraude au clic (clic fraud).

V.2.g La fraude au clic.

La fraude au clic consiste à cliquer massivement sur les liens sponsorisés pour augmenter artificiellement la facture dont des entreprises concurrentes, annonceurs sur Google, devront s'acquitter. Clic Defense⁷⁵, qui propose des solutions pour lutter contre cette pratique, estime que 34% des clics sur les liens sponsorisés sont frauduleux.

Clic Defense accuse le moteur de recherche de n'avoir pas suffisamment informé et protégé ses clients des dangers des liens sponsorisés, ce qui leur aurait fait perdre plus de 5 millions de dollars⁷⁶.

Or, Click Defense a acheté des mots clés à Google afin d'assurer la visibilité des sites Web de ses clients sur le moteur le plus visité du moment.

L'agence Web marketing estime que les budgets consommés en pure perte s'établissent à 5 millions de dollars.

Click Defense a déposé plainte le 24 juin 2005 devant la cour de district de San Jose, Californie. L'agence réclame des dommages et intérêts à Google, et souhaite obtenir le statut de "class action" (action collective).

⁷⁵ spécialiste des outils marketing sur Internet

⁷⁶ Source : Reuters jeudi 30 juin 2005, repris par <http://www.lexpansion.com/html/A133169.html>

Google a nié les allégations trouvant que la plainte était « sans fondement ».

Conséquences de ces business models sur le monde de l'information de la connaissance et des langues

L'introduction de nouveaux modèles d'affaires dans l'industrie des outils de recherche redéfinit les règles du traitement des résultats renvoyés par les moteurs de recherche. De quel degré d'information disposent les internautes face à l'ingérence du commerce dans leurs surfs quotidiens et comment réagissent-ils face à cette cohabitation entre moteurs de recherche et commerce électronique ?

Nous verrons également de quelle manière les utilisateurs peuvent eux-mêmes être des sources d'information, souvent bien malgré eux.

Nous apporterons des recommandations afin d'améliorer la protection du consommateur sur internet et tenterons de lui proposer des outils méthodologiques pour améliorer la recherche d'information. Nous nous interrogerons ensuite sur la répartition des langues sur le Web, facteur non négligeable d'accès à l'information mais aussi de diffusion de cette dernière. L'information multilingue, et a fortiori le web multilingue ne sont-ils que des fantasmes ?

VI. Conséquences de ces business models sur le monde de l'information de la connaissance et des langues

VI.1 Le problème de la pertinence de l'information sur Internet

Nous avons bien vu que, contrairement à l'idée d'une bibliothèque d'Alexandrie, Internet n'avait rien d'une bibliothèque, ni d'une cyber-encyclopédie et nous savons bien que les informations disponibles sur la toile ne sont ni sélectionnées, ni organisées.

L'accès aux informations pertinentes dépend donc, pour l'utilisateur, de :

- la capacité à utiliser la syntaxe des outils de recherche (Les interfaces et les fonctionnalités avancées des moteurs de recherche sont comparables avec les interfaces thématiques et multicritères de BCDI)
- la capacité à choisir un outil de recherche adapté (annuaire, moteur, métamoteurs)
- la capacité à formuler des mots clés et des expressions clés susceptibles de se trouver dans le texte du document que l'on souhaite trouver.
- la capacité d'identifier les sources d'information et de juger de leur pertinence, de leur fiabilité et de leur exactitude.
- La capacité à analyser le fonctionnement des outils de recherche eux-mêmes (méthodes d'indexation des moteurs de recherche, critères de classement utilisés pour l'affichage des résultats, techniques de référencement) est elle aussi indispensable.

De nombreux travaux théoriques et expérimentaux ont eu pour but de définir les caractéristiques de la pertinence en recherche d'information [Saracevic, 1996], [Schamber, 1994]. Les recherches sur la définition de la pertinence sont anciennes, puisqu'elles datent des années 50. [Le Coadic, 1998] présente et synthétise une chronologie de ces recherches du début des années 1950 jusqu'à l'année 1992. Comme, il le constate, on peut regrouper ces études en deux groupes. Le premier groupe renvoie à la notion de pertinence objective (point de vue du système) et le second groupe rassemble les notions de pertinence subjective (point de vue de l'utilisateur).

Pour les tenants du paradigme système, un "bon" système est celui qui trouve tous les documents pertinents avec un minimum de références non pertinentes. Plusieurs auteurs dont [Saracevic, 1996] et [Shamber, 1994] ont montré que les critères de performance (rappel/précision) sont insuffisants pour déterminer la pertinence d'une recherche.

Dans l'approche traditionnelle on présume que les jugements de pertinence :

- sont des indicateurs d'efficacité des SRI
- sont stables et ne varient pas dans le temps
- peuvent être réduits à un jugement binaire : un document est soit pertinent, soit non pertinent. Or, il est difficile de fixer un seuil de pertinence à partir duquel un document est pertinent.
- sont attribués indépendamment les uns des autres. Les jugements de pertinence doivent donc être nuancés. La pertinence d'un document est relative à l'ensemble des documents consultés.

Les usagers utilisent des critères autres que ceux du rappel et de la précision qui restent étrangers à leurs préoccupations, lorsqu'ils initient ou terminent une session de recherches.

Les usagers effectuant la même requête et se trouvant confrontés aux mêmes ensembles de documents, peuvent avoir des avis différents sur la pertinence des documents. Leur jugement tient compte en effet de plusieurs facteurs liés au contexte de recherche (leur niveau de connaissances, leur expérience du domaine, etc.) et à la nature de leur besoin en information.

Une grande diversité de facteurs entre en jeu lorsqu'un individu évalue la pertinence d'un document. Pour [Schamber, 1994], la pertinence est subjective, cognitive et multidimensionnelle. L'auteur a identifié quatre vingt facteurs qui agissent ou influent sur la pertinence d'un document.

Il existe plusieurs facteurs qui interviennent dans l'évaluation de la pertinence des réponses.

On ne peut pas parler d'un seuil de pertinence d'un document mais d'un système de pertinence de l'utilisateur. [Mucchielli, 1995] notait à juste titre que le système de pertinence d'un individu *"est un état psychologique de prédisposition mettant en cause le cognitif, l'affectif, le perceptif et le comportemental. Il est en fonction de l'ensemble des problèmes spécifiques qui préoccupent l'individu, des projets qu'il a et qui forment son orientation de vie au moment où on le considère"*. Il y a donc une relation forte entre le projet d'un usager, ses buts et une perception sélective des facteurs qui peuvent agir sur son système de pertinence.

VI.2 État de conscience des consommateurs (utilisateurs des moteurs de recherche)⁷⁷

Les chercheurs de Consumer Reports WebWatch ont réunis dans le cadre d'une étude ethnographique dont les résultats sont parus en juin 2003, un groupe hétérogène d'internautes. Bien que chaque participants ait eu au minimum cinq ans d'expérience en matière d'utilisation de l'Internet, aucun n'avait une compréhension bien définie concernant le fonctionnement des moteurs de recherche et leurs méthodes de classement des résultats. Ils ne savaient pas non plus de quelle manière l'inclusion de résultats de recherche achetés aux enchères par les partenaires commerciaux des moteurs de recherche pouvait influencer les premières pages de résultats qu'ils visualisaient sur leurs écrans. De nombreux participants, ne réalisaient pas, par exemple, que les moteurs de recherche étaient des entreprises à but lucratif. Il n'était venu à l'idée d'aucun d'entre eux que le classement des résultats qu'ils recevaient lors de leurs recherches pouvait subir l'influence d'un commerce bilatéral voire multilatéral entre la société du moteur de recherche et d'autres sociétés. Jusqu'à ce que ce fait soit mis en évidence par un chercheur en charge de l'expérience.

Ces erreurs de jugement concernant les moteurs de recherche nous laissent penser que les utilisateurs les moins expérimentés doivent être d'autant moins au fait du commerce lié à la recherche d'information et au référencement payant. Les découvertes faites lors de cette étude montre que les internautes choisissent une fois sur deux de cliquer sur un résultat affiché dans la première page de résultats car ils « font confiance » au moteur pour leur présenter en priorité les résultats les plus pertinents. Cette même « confiance » fait qu'ils jugent inutile le fait de consulter les pages suivantes.

Cette mentalité des « premiers résultats sont les meilleurs » rend cette population plus susceptible de cliquer sur les liens commerciaux des pages de résultats, en particulier si ces liens ne sont pas clairement signalisés afin de les différencier de résultats « naturels ».

Cette confusion dans l'esprit des utilisateurs est entretenue par l'utilisation de notions vagues ou peu explicites utilisés sur les moteurs de recherche telles que « sponsorisé », qui selon la majorité des participants est un terme trompeur.

Cette facilité à cliquer sur les liens commerciaux rend ces internautes plus vulnérables car ils sont plus exposés à une information erronée ou biaisée émanant de sites Web d'entreprises ayant payé pour figurer dans les pages de résultats, sans pour autant disposer d'une information répondant précisément à la requête des internautes. Ce biais peut induire des conséquences plus graves qu'une simple perte de temps si, par exemple, la recherche est strictement informationnelle, disons dans le cadre d'informations sur un traitement de maladie, ce simple clic sur un lien commercial peut avoir un impact sur la santé de l'internaute. Dans ce type de situation, il est

⁷⁷ D'après les résultats d'une étude menée par Leslie Marable, chercheur/écrivain, pour Consumer Reports WebWatch en juin 2003.

possible qu'un résultat ayant un moins bon classement aurait pu répondre à l'objet de la requête de façon plus pertinente.

De plus, les participants n'ont pas remarqué les liens d'information ou les liens « About Search⁷⁸ » sur les quinze sites soumis à l'essai, jusqu'à ce qu'un ethnologue ne les leur montre. La plupart des participants ont alors dit que ces liens informatifs étaient trop petits pour que les internautes puissent les voir, ou qu'il étaient situés à des endroits de la page où il était peu commode de regarder. Le problème est d'autant plus important lorsqu'il s'agit de métamoteurs, où les liens informatifs apparaissant sur les moteurs d'origine disparaissent parfois lorsque les résultats sont compilés par les filtres et l'algorithme de classement du métamoteur ?

Les participants ont donc été surpris lorsqu'ils ont appris le fonctionnement du référencement payant des moteurs les plus populaires :

De nombreux participants ont eu une réaction négative, allant de la colère et du sentiment de tromperie, jusqu'à un sentiment d'impuissance et de profonde déception. La plupart, poursuit l'étude, ont exprimé une confiance altérée dans les moteurs de recherche tandis que certains se sont montrés plus suspicieux envers les annonceurs dont les résultats « achetés » s'étaient montrés non pertinents.

Ces internautes sont ensuite devenus plus critiques dans leur traitement des résultats et ils se sont mis à développer de nouvelles méthodes de visualisation et d'exploitation des pages de résultats, avant de prendre la décision de cliquer ou non sur un lien : par exemple, ils parcourent un plus grand nombre de pages de résultats et font l'effort de consulter en priorité les résultats « naturels ». Ces pratiques se sont révélées d'autant plus systématiques lorsque l'objet de la recherche était strictement informationnel, lorsque l'utilisateur effectue une recherche d'information sans intention d'achat et qu'il ne souhaite pas voir de liens commerciaux.

Les chercheurs ont demandé aux participants de leur dire de quelle manière, selon eux, les sociétés de moteurs de recherche pourraient rendre les résultats commerciaux plus faciles à repérer. Presque tous ont suggéré une dénomination claire et non ambiguë telle que « liens publicitaire », ils ont aussi soutenue l'idée d'une signalétique particulière qui aiderait les utilisateurs à repérer rapidement les zones des pages de résultats correspondant à des liens naturels ou commerciaux, comme par exemple l'usage de cadres à fond colorés et de zones exclusivement réservées aux liens commerciaux.

Les participants ont exprimés un regret concernant le fait que l'industrie de la recherche d'information en ligne n'ait pas décidé d'un commun accord d'un vocabulaire et d'une signalétique qui serait identique d'un moteur de recherche à l'autre, en appelant à la formation d'un comité de réflexion constitué de professionnels de l'industrie et à l'intervention du gouvernement américain sur ces questions.

Pour autant, les utilisateurs des moteurs de recherche ne considèrent pas le référencement payant comme quelque chose de « mauvais », la plupart confient même

⁷⁸ « Au sujet de la recherche »

que l'utilisation de pages de liens commerciaux dans le cadre d'une recherche motivée par une intention d'achat représente un outil efficace qui les aide à trouver les sites commerciaux disposant des produits et services qu'ils recherchent.

Malgré cela, chaque utilisateur a le droit de connaître à l'avance la nature du lien sur lequel il est susceptible de cliquer. Eduquer l'internaute sur ces pratiques lui permet d'améliorer sa prise de décision. En effet, l'utilisateur a le droit de savoir qu'une information qui serait plus pertinente pour sa recherche d'information a toutes les chances de se retrouver dans le bas du tableau, simplement parce que la page Web sur laquelle elle se trouve est la propriété d'une petite entreprise ou d'une organisation à but non lucratif qui ne dispose pas des moyens suffisants pour acheter sa place au sommet des premières pages de résultats. Les résultats payants prennent quelque peu d'assaut, en l'état actuel des choses, les internautes, un peu comme si on interrompait chaque reportage du journal télévisé avec une publicité traitant vaguement du sujet. Les consommateurs demandent simplement à l'industrie de regrouper les publicités, comme à la télévision, dans des pages de publicités dédiées, leur offrant ainsi la possibilité de zapper, car c'est la confiance des utilisateurs en l'industrie des outils de recherche qui est en jeu.

VI.3 Les informations personnelles/ La CNIL

Il existe aujourd'hui, à la suite du développement du commerce électronique, une situation nouvelle et inconfortable pour les usagers. Les entreprises présentes sur les réseaux cherchent à extraire des informations personnelles en provenance de leurs utilisateurs. Ils constituent des bases de données d'informations personnalisées afin de toucher plus directement les consommateurs grâce au marketing ciblé (" one to one marketing "). On peut aujourd'hui suivre le profil d'un utilisateur au cours du temps, connaître les sites visités avant de parvenir sur celui de l'entreprise, réaliser des statistiques sur les comportements de telle ou telle famille. Ces informations mettent en danger la vie privée des utilisateurs. Ceux-ci commencent d'ailleurs à s'en rendre compte et cherchent à protéger par des moyens appropriés, existant d'ailleurs sur les réseaux, les informations qu'ils considèrent comme personnelles et non utilisables dans la sphère publique.

VI.3.a La CNIL

Une méfiance est donc en train de naître entre les usagers des réseaux et les entreprises de biens et de services. C'est pourquoi va apparaître une nouvelle fonction d'intermédiaires entre les producteurs et les consommateurs, une sorte de tiers de confiance, devenu obligatoire pour assurer les consommateurs de la garantie d'utilisation des informations privées qu'ils font circuler sur les réseaux : la CNIL.

- **L'institution**

La Commission Nationale de l'Informatique et des Libertés (CNIL) a été instituée par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés qui la qualifie d'autorité administrative indépendante.

La CNIL c'est :

Un collège pluraliste de 17 commissaires :

- * 4 parlementaires (2 députés, 2 sénateurs),
- * 2 membres du Conseil économique et social,
- * 6 représentants des hautes juridictions (2 conseillers d'État, 2 conseillers à la Cour de cassation, 2 conseillers à la Cour des comptes),
- * 5 personnalités qualifiées désignées par le Président de l'Assemblée nationale (1 personnalité), par le Président du Sénat (1 personnalité), par le conseil des ministres (3 personnalités).

Le mandat de ses membres est de 5 ans ou, pour les parlementaires, d'une durée égale à leur mandat électif.

Pour conduire leurs missions, les membres de la CNIL s'appuient sur différents services.

Une autorité indépendante :

12 des 17 membres sont élus par les assemblées ou les juridictions auxquelles ils appartiennent.

La CNIL élit son Président parmi ses membres ; elle ne reçoit d'instruction d'aucune autorité ; les ministres, autorités publiques, dirigeants d'entreprises, publiques ou privées, ne peuvent s'opposer à l'action de la CNIL pour quelque motif que ce soit et doivent prendre toutes mesures utiles afin de faciliter sa tâche.

Le Président de la CNIL recrute librement ses collaborateurs.

Une autorité administrative :

- * Le budget de la CNIL est imputé sur le budget de l'État.
- * Les agents de la CNIL sont des agents contractuels de l'État.
- * Les décisions de la CNIL peuvent faire l'objet de recours devant la juridiction administrative.

Le travail de la CNIL se justifie relativement facilement et une simple démonstration sur le site de la CNIL nous permet d'en prendre rapidement conscience.

A travers notre adresse IP, le Proxy que l'on utilise, notre navigateur, c'est une vraie petite carte d'identité en miniature que chaque site peut bâtir de nous et que chaque système d'intelligence marketing peut exploiter pour le compte d'une entreprise à des fins commerciales. Voyons les principales informations que nous laissons sur le Web plus ou moins involontairement. N'oublions pas que le disque dur de notre ordinateur contient lui aussi des informations capitales nous concernant, et auxquelles peuvent avoir accès les internautes dès lors qu'on leur en offre l'accès par le biais d'un logiciel de peer-to-peer (partage de fichiers).

Face aux dangers que l'informatique peut faire peser sur les libertés, la CNIL a pour mission essentielle de protéger la vie privée et les libertés individuelles ou publiques. Elle est chargée de veiller au respect de la loi "Informatique et Libertés" qui lui confie 5 missions principales :

- **Informer**

La CNIL informe les personnes de leurs droits et obligations, et propose au gouvernement les mesures législatives ou réglementaires de nature à adapter la protection des libertés et de la vie privée à l'évolution des techniques. L'avis de la CNIL doit d'ailleurs être sollicité avant toute transmission au Parlement d'un projet de loi créant un traitement automatisé de données nominatives.

- **Garantir le droit d'accès.**

La CNIL veille à ce que les modalités de mise en oeuvre du droit d'accès aux données contenues dans les traitements n'entravent pas le libre exercice de ce droit. Elle exerce, pour le compte des citoyens qui le souhaitent, l'accès aux fichiers intéressant la sûreté de l'État, la défense et la sécurité publique, notamment ceux des Renseignements généraux.

- **Recenser les fichiers.**

Les traitements de données à "risques" sont soumis à autorisation de la CNIL. Elle donne un avis sur les traitements publics utilisant le numéro national d'identification des personnes. Elle reçoit les déclarations des autres traitements. Le non-respect de ces formalités par les responsables de fichiers est passible de sanctions administratives ou pénales. La CNIL tient à la disposition du public le "fichier des fichiers", c'est-à-dire la liste des traitements déclarés et leurs principales caractéristiques.

- **Contrôler**

La CNIL vérifie que la loi est respectée en contrôlant les applications informatiques. La Commission use de ses pouvoirs de vérification et d'investigation pour instruire les plaintes, pour disposer d'une meilleure connaissance de certains fichiers, pour mieux apprécier les conséquences du recours à l'informatique dans certains secteurs, pour assurer un suivi de ses délibérations. La CNIL surveille par ailleurs la sécurité des systèmes d'information en s'assurant que toutes les précautions sont prises pour empêcher que les données ne soient déformées ou communiquées à des personnes non autorisées.

La CNIL peut prononcer diverses sanctions graduées : avertissement, mise en demeure, sanctions pécuniaires pouvant atteindre 300 000 € injonction de cesser le traitement. Enfin, le Président peut demander par référé à la juridiction compétente d'ordonner toute mesure de sécurité nécessaire. Il peut, au nom de la Commission, dénoncer au Procureur de la République les violations de la loi.

- **Réglementer**

La CNIL établit des normes simplifiées, afin que les traitements les plus courants et les moins dangereux pour les libertés fassent l'objet de formalités allégées.

Elle peut aussi décider de dispenser de toute déclaration des catégories de traitement sans risques.

VI.3.b Nos informations personnelles

- **Le numéro IP.**⁷⁹

Sur un réseau TCP/IP tel Internet, le numéro IP identifie chaque ordinateur connecté au réseau. Un ordinateur connecté en permanence à Internet conservera toujours le même numéro IP, mais, pour la majorité des particuliers, l'adresse Ip est dynamique, c'est-à-dire qu'elle change à chaque connexion.

Grâce à votre numéro IP, on peut déterminer votre fournisseur d'accès Internet et vous identifier

L'adresse IP est codée sur 32 bits. Elle comporte deux champs : un identifiant réseau et un identifiant machine.

Face à l'accroissement du nombre d'ordinateurs (il faut que chaque matériel connecté ait une adresse IP distincte), une norme IPv6 codant l'adresse sur 128 bits au lieu de 32 a été mise en place. IPv6 conserve la compatibilité avec les adresses IP classiques. Répondre à la croissance exponentielle du nombre d'ordinateurs connectés à Internet n'est pas le seul enjeu d'IPv6. En plus d'une extension de l'adressage, IPv6 est censé permettre une meilleure qualité de service car cette norme définit des niveaux de priorité lors du transport des données. Ce qui permettra d'ajuster la qualité du service aux besoins de l'application et d'envisager de nouvelles applications (temps réel, multimédia,...). La mise en place d'IPv6 passe par la mise en place progressive de nouveaux routeurs.

Si vous n'utilisez pas de Proxy (pour le savoir, regarder dans les options de votre navigateur), tous les sites Web consultés connaissent votre adresse IP (c'est votre ordinateur qui demande sans intermédiaire le contenu des pages que vous souhaitez consulter) car elle est incluse dans chaque paquet de données qu'envoie votre ordinateur sur le réseau. Rien n'empêche un site de créer un fichier journal des demandes d'information contenant votre adresse IP, la réponse faite par le serveur du site,...

Si vous utilisez un Proxy, c'est le Proxy qui demande les données pour vous et vous les renvoie. Le site consulté ne connaît donc que l'adresse IP de votre proxy. Attention, c'est souvent le Proxy de votre fournisseur d'accès. Cela n'augmente donc pas vraiment votre "anonymat" car le Proxy de votre FAI pourra très bien conserver l'intégralité de l'historique de vos requêtes ! Un site client pourra donc, en s'adressant à votre FAI (déterminé à partir de l'adresse IP du Proxy), remonter jusqu'à vous. Mais il existe des proxys anonymes c'est-à-dire qu'ils ne tiennent pas de "journal" et cachent bien sûr comme tous les proxys votre adresse IP réelle. L'utilisation d'un tel Proxy ne se traduit aucunement par une diminution de la qualité de connexion. Pour ceux qui seraient intéressés par une telle manipulation, voici une liste indicative de proxys "anonymes".

Sous Windows, vous pouvez savoir quelle est votre adresse IP lors d'une connexion à Internet en démarrant le programme winipcfg (Démarrer \ Exécuter).

⁷⁹ Source yann.fauche.free.fr

- **Les cookies.**

"Les cookies sont des informations engendrées par le serveur Web consulté et stockées sur le disque des utilisateurs, à leur insu, lors de chaque visite. Ces informations seront exploitées ultérieurement par le serveur au cours des prochaines connexions."⁸⁰

Les cookies sont donc des bribes d'information que déposent sur notre disque dur de nombreux sites Web dans le but de nous "reconnaître"(reconnaître l'ordinateur) **lors de notre prochaine visite.** Avec ce "lien", au fur et à mesure de nos visites successives sur un site, le site peut construire une sorte de profil de vous.

Cela peut être une bonne chose comme une mauvaise, tout dépend de l'usage fait par le site de l'information récoltée.

"Voici quelques exemples d'utilisations possibles des cookies :

- un serveur d'actualité ou d'articles de presse vous demande de remplir un formulaire pour indiquer vos préférences (sports : football, politique : écologie, spectacle : cinéma), afin de les stocker dans un cookie sur votre PC pour vous présenter directement, lors de vos prochaines connexions, les informations correspondant à ces préférences.
- un serveur de commerce électronique insère un cookie à chaque fois que l'utilisateur sélectionne un produit, au même rythme que celui du remplissage d'un caddie. Lorsque l'utilisateur se rendra sur la page contenant le formulaire de commande, le serveur récupérera l'ensemble des cookies du "caddie" afin d'afficher l'ensemble des produits sélectionnés par le client.
- un moteur de recherche de site positionne des cookies en fonction des rubriques visitées par le client (Ex. : informatique, musique, santé) afin, ultérieurement, d'afficher dynamiquement des bandeaux publicitaires correspondant aux goûts ainsi décelés.
- un serveur propose à l'utilisateur de choisir lui-même la couleur du fond d'écran, la présence de multi-fenêtrage ou encore les polices de caractère utilisées. Il stocke toutes ces informations dans des cookies et présente au client une page d'accueil correspondant précisément à ses goûts graphiques."⁸¹

On ne peut donc nier le caractère utile des cookies lors de notre navigation sur le Web. Mais le risque que l'information récoltée au fur et à mesure de nos visites soit vendue, échangé ou regroupé n'est pas nul.

Certains cookies sont plus pernecieux que d'autres. Seul le site qui dépose qui dépose un cookie est capable de le relire dans le futur mais il arrive qu'en visitant un site (A) quelconque ce soit un autre site (B) qui dépose un cookie sur votre ordinateur. C'est courant. Il suffit que sur le site (A) il y est un lien vers une image (souvent de la publicité) du site (B). L'image sera donc téléchargée sur le serveur de (B) et (B) pourra stocker un cookie sur votre ordinateur ! Elle s'affichera comme tout le reste sur le site (A) mais le site (B) vous aura (généralement dans un but publicitaire) "marqué". Supposons que le site (B) arrive à inclure son lien dans de nombreux sites. C'est un profil complet de vous qu'il sera en mesure de bâtir (les pages que vous visitez, l'ensemble des informations que vous avez donné,...).

⁸⁰ Source : encyclopédie Universalis

⁸¹ extrait de la documentation CNIL

Ces cookies qui sont stockés par un site autre que celui que vous pensez être en train de visiter sont appelés "foreign cookies". Une autre utilisation plutôt "immorale" des cookies est faite par certains sites de vente en ligne. Lorsque vous avez acheté un produit en ligne, cela est marqué sur un cookie stocké sur votre ordinateur et, lors d'un prochain achat, le prix du nouveau produit acheté sera fonction des achats déjà effectués. Ce procédé peut bien sûr servir à la mise en place de réduction pour les clients les plus fidèles mais, plus couramment, il permet de faire payer le prix fort à ceux qui sont moins vigilants ! Vous ne bénéficierez de tarifs attractifs que lors du premier achat sur ce site.

Un cookie n'est pas un programme : il ne peut pas contenir de virus ou avoir une action quelconque. C'est simplement un fichier texte (donc que vous pouvez ouvrir et modifier). Il stocke seulement des informations sur vous mais en aucun cas ne recherche quoi que ce soit sur votre disque dur. C'est grâce à l'aspect ouvert d'Internet que l'existence des cookies est connue. Les cookies ne sont pas des programmes-espions.

C'est votre navigateur (par exemple Internet Explorer, Netscape Navigator,...) qui gère les cookies. Lorsqu'un site veut mettre en place un cookie, il en demande l'autorisation à votre navigateur. Vous pouvez donc généralement configurer votre comportement face aux cookies dans les options de votre navigateur Web⁸².

En conclusion, si les cookies présentent parfois des avantages évidents pour le netsurfeur, il présente aussi un aspect camouflé, dissimulé, et dangereux pour notre vie privée.

Il semble difficile de refuser complètement l'utilisation des cookies. Ce serait refuser le confort qu'ils peuvent nous offrir et s'interdire l'utilisation de certains sites multimédia dont le fonctionnement repose sur leur utilisation. Il n'y a pas de solution parfaite.

- **Les variables d'environnement de votre navigateur.**

En plus de votre adresse IP, chaque demande de consultation d'une page Web se traduit par l'envoi d'une certaine quantité d'informations supplémentaires : à chaque paquet de données que vous envoyez au serveur se trouve adjoint des informations qui permettent au site que vous consultez de déterminer votre navigateur, généralement votre système d'exploitation, et aussi le lien que vous avez suivi pour atterrir sur la page demandée. Ces données sont traitées par le serveur sous le nom de variables d'environnement :

REMOTE_HOST =	<i>le site client (votre adresse DNS)</i>
REMOTE_ADDR =	<i>votre adresse IP</i>
HTTP_USER_AGENT =	<i>la signature de votre navigateur</i>
HTTP_REFERER =	<i>la page précédente</i>

⁸² Avec le navigateur Internet Explorer, les cookies se trouvent dans le répertoire "C:/WINDOWS/Cookies".

Les variables d'environnement sont un outil intéressant pour les sites Web. Elles permettent par exemple d'adapter la page affichée au navigateur utilisé. Le danger des variables d'environnement réside dans leur association avec les autres sources d'information du site (cookies,...).

- **Javascript :**

Lorsque que vous consultez une page, elle peut contenir du code Javascript. Ce code, qui s'exécute sur votre machine, peut renvoyer au serveur des informations : nombre de page consultés durant cette session (mais pas leur adresse), taille de la fenêtre de votre navigateur, compatibilité java, plugins installés sur votre navigateur...

La recherche de l'anonymat et la protection de notre vie privée sur Internet ne sont donc pas des choses évidentes à garantir. Elles nécessitent comme dans la vie courante une réelle volonté et motivation de l'utilisateur. Il est important que l'ensemble des acteurs du réseau Internet prenne conscience des enjeux et des risques des technologies qu'ils utilisent!

Le maillon faible dans la protection de notre vie privée reste notre fournisseur d'accès Internet. De part sa place privilégiée, il est le plus à même de récolter facilement des informations sur nous. "S'est-il engagé à ne pas analyser nos connexions ? Sinon, s'est-il engagé à utiliser ces informations pour des finalités précises et limitées ? S'est-il engagé à ne pas transmettre ces informations à des tiers ? S'est-il engagé sur une durée de conservation de ces informations ?"⁸³

C'est uniquement à partir de notre démarche responsable et engagée à tous que pourra se bâtir le respect de la vie privée. Internet n'est pas un espace "hors-frontière". La même logique de transparence demandée aux moteurs de recherche quant à l'usage qu'ils font des mots-clés pour gagner de l'argent doit être réclamée aux entreprises et aux administrateurs de sites Web pour qu'ils nous fournissent des informations sur l'usage qui est fait des informations privées que nous laissons. Il n'est à cet égard pas inutile de rappeler que collecter des informations nominatives est une démarche responsable dans laquelle l'entreprise (le site) s'engage vis à vis des utilisateurs.⁸⁴

En 2005, Reporters sans frontières (RSF) a révélé que le géant américain de l'Internet Yahoo ! a remis à la justice chinoise des informations qui ont permis d'envoyer un journaliste chinois en prison pour dix ans. Jerry Yang, le cofondateur de Yahoo, a expliqué qu'il n'avait pas le choix et que sa société était contrainte d'obéir aux lois des pays dans lesquels elle opérait. La réputation de Yahoo risque d'en être durablement entachée et son site est menacé de faire l'objet de campagnes de boycottage. Shi Tao, un journaliste de 36 ans, employé du journal Dangdai Shang Bao (Contemporary Trade News) de Changsha, la capitale de la province du Hunan, avait adressé un e-mail, le 20 avril 2004, à un site Internet de langue chinoise hébergé à l'étranger. Il y révélait le contenu d'une note confidentielle reçue par son journal et concernant le climat politique à la veille du quinzième anniversaire du massacre de Tiananmen, le 4 juin. Le département de la propagande du Parti communiste chinois donnait ses consignes à la presse officielle face aux «risques pour la stabilité». Arrêté en novembre 2004, Shi Tao est condamné le 30 avril 2005 à dix ans de prison pour

⁸³ Source : documentation CNIL

⁸⁴ Source : <http://yann.fauche.free.fr>

«divulgarion de secrets d'Etat à l'étranger», une peine très lourde au regard des faits. Le déroulement du procès avait déjà été critiqué à l'étranger : à la veille de l'audience, l'avocat du journaliste, Guo Guoting, a été privé du droit d'exercer pendant un an et n'a donc pas assuré la défense... L'affaire en serait restée là, c'est-à-dire sans faire trop de vagues, si RSF n'avait pas révélé que la preuve décisive du procès a été fournie par Yahoo. Le message envoyé par Shi Tao était anonyme, mais, à la demande de la justice chinoise, Yahoo a révélé l'identité du détenteur de l'adresse IP qui s'y rattache, en quelque sorte l'ADN du message électronique, et le numéro de téléphone utilisé pour l'expédier, celui des bureaux du journal⁸⁵. Ce fut donc le procès de l'éthique contre la loi nationale et une remise en question du Web, « espace de liberté » !

⁸⁵ Source : Libération, 16 septembre 2005

VI.4 Comment adapter le modèle et protéger le consommateur?

Les moteurs de recherche doivent être plus transparents dans leur communication avec les utilisateurs.

Par exemple, ils pourraient fournir une liste exhaustive de leurs partenaires commerciaux et préciser les relations commerciales susceptibles d'influencer d'une quelconque manière le classement des résultats d'une recherche.

Consumer Webwatch propose d'afficher cette liste de façon bien visible et de préciser la date courant ainsi qu'une date de dernière mise à jour. Les meilleurs sites pourraient relayer cette information sur chacune des pages de résultats ou offrir la possibilité d'accéder à cette liste via un lien bien visible lui-même inscrit sur chaque page.

Ensuite, les sites devraient fournir à l'internaute des explications claires sur leur algorithme de classement des résultats, en vulgarisant les notions techniques telles que Webspider, crawler, indexation manuelle, etc.

Il serait aussi utile d'expliquer de quelle manière les résultats achetés sont pondérés par rapport aux résultats naturels avant d'être classés sur la page de résultats.

Les moteurs pourraient signaler aux utilisateurs si les résultats commerciaux sont exclusifs à leur site, et dire également quels résultats appartenant à ce même partenaire commercial sont générés de façon « naturelle », et d'expliquer les appellations "Enhanced by Search Engine X" ou "Powered by Search Engine Y."

Les meilleurs sites devront utiliser une terminologie claire et sans ambiguïté pour désigner les résultats commerciaux comme « liens promotionnel acheté par X » toujours dans le but d'améliorer la transparence envers l'utilisateur et de lui permettre de distinguer les résultats naturels.

A l'instar de Yahoo ! Mindset, les moteurs devraient proposer à l'utilisateur de configurer sa recherche en fonction de son intention commerciale ou informationnelle et ce, le plus tôt possible dans le processus de recherche.

Les meilleurs sites doivent expliquer aux utilisateurs le biais introduit par la théorie du « le premier résultat sur la liste est toujours le meilleur ».

Les sites devraient fournir un lexique contenant les définitions de termes marketing de base tels que « coût par clic » ou « liens promotionnels », « positionnement payant », « pay per click », « paid inclusion » et « soumission payante ».

Les sites offrant des services de keyword marketing pour les annonceurs devraient éviter d'employer des expressions telles que « le meilleur » « the best » « most relevant » ou « most matches ».

En ce qui concerne les métamoteurs, il faudrait que soient notifiés les résultats qui constituent des « liens publicitaires ».

L'usage de masse des outils de recherche pose également de nombreux (et relativement nouveaux) problèmes, sur les modifications des pratiques informationnelles (comme le "zapping"), sur les nouveaux modes de connaissance induits par les logiques des outils de recherche, sur les dangers de "l'info-pollution", etc.

Une première observation très simple frappe l'observateur, devant le paysage actuel des outils et des méthodes de recherche : nous sommes passés, en deux décennies à peine, d'une relative pauvreté, ou rareté des outils, à une abondance, une explosion, une prolifération d'outils de recherche ; dans les années 80, par exemple, la recherche d'information se limitait aux logiciels documentaires (environ une vingtaine), aux logiciels des serveurs de banques de données, et à quelques outils documentaires spécifiques. Aujourd'hui, les outils de recherche, uniquement sur Internet, se comptent par milliers, il en meurt et il en naît chaque année plusieurs dizaines et la diversité, dans la spécialisation, est infinie.

Deuxième observation : l'imbrication, l'hybridation des modes de recherche et des outils. On peut distinguer, schématiquement, quatre modalités de recherche d'information : la navigation arborescente (dans les annuaires thématiques, les classifications), la navigation hypertextuelle (dans les sites Web, les CD-ROM, les encyclopédies), la recherche par requête sur des mots-clés dans des champs délimités (l'interrogation des banques de données, des catalogues...) et la recherche par requête sur le contenu (recherche en texte intégral, moteurs de recherche). A chacune de ces modalités correspondaient des pratiques, des usages de recherche, mais aussi des outils, jusqu'alors bien distincts.

Ainsi, aux débuts du Web, la distinction entre annuaires et moteurs était-elle parfaitement claire, du seul point de vue de la modalité de recherche.

Or l'une des évolutions profondes de la "RI" a consisté à entremêler ces diverses modalités ainsi que les outils sur lesquels elles s'appuient. Depuis quelques années, la mixité entre annuaires et moteurs (combinant recherche arborescente et sur le contenu), ou le développement des portails, proposant tous les types de recherche, témoignent de cette imbrication d'outils, de techniques et de modalités de recherche différents, ajoutant parfois de la confusion au paysage et rendant les distinctions de plus en plus difficiles. Pour autant, les typologies d'outils restent fondamentales à maîtriser et à expliquer aux utilisateurs, pour sortir des "apparences de l'écran".

Tout d'abord, l'utilisateur du moteur de recherche doit avoir conscience, qu'à l'instar de tout autre commerce, les sociétés de services qui offrent un accès à leur moteur de recherche ont besoin de gagner de l'argent pour exister.

En effet, le coût engendré par l'indexation de milliards de pages chaque mois est très important et les sociétés de moteurs de recherche qui fournissent des services de recherche fondés sur des algorithmes de mesure de pertinence telles que Google, AskJeeves, doivent obtenir des revenus pour soutenir leurs opérations quotidiennes, d'autant plus qu'elle doivent affronter la concurrence.

Pour cette raison, il ne faut pas diaboliser le keyword marketing. Il ne faut pas tomber dans un excès inverse qui supposerait que, parce qu'un résultat a été acheté aux enchères, il n'est pas pertinent pour la requête formulée. Le site concerné peut très bien proposer l'information juste. De la même manière, il ne faut pas considérer que tous les résultats affichés ont été achetés par des entreprises.

Plusieurs sociétés de moteurs de recherche possèdent des moteurs divers ainsi, InfoSpace possède InfoSpace.com, Dogpile.com, Metacrawler.com... ce qui explique que l'on puisse retrouver les mêmes résultats de recherche sur chaque moteur appartenant à la même marque ombrelle.

VI.5 L'importance d'une démarche pédagogique⁸⁶ :

Apprendre les méthodes d'interrogation de chaque outil est importante mais comme ils sont appelés à se modifier et que de nouveaux apparaissent, l'internaute doit surtout comprendre la logique d'une recherche et ses grands principes, qui eux restent les mêmes, pour faire face à la nouveauté.

Un des sites les plus éclairants est celui des cours offerts par Dykes Library (<http://www.kumc.edu/service/dykes/classes/search.html>).

Leur introduction explique la problématique de la recherche d'information, en quoi l'Internet diffère des banques de données commerciales traditionnelles, et pourquoi il est important pour le succès d'une recherche de comprendre comment les logiciels robots fonctionnent.

Une recherche fine passe ainsi par la compréhension du mode d'interrogation des différents outils et la connaissance de la syntaxe propre à chaque moteur utilisé.

VI.5.a Le constat:

Dans la majorité des cas où l'utilisateur doit demander une information à un documentaliste, sa requête ressemblera à :

« Qu'est-ce que vous avez sur... », ou « Trouvez-moi tout sur ... ».

Cette formulation occulte généralement les éléments nécessaires à la recherche d'informations :

- la question réelle
- le contexte
- la problématique

VI.5.b Les types de question auxquelles l'utilisateur peut être confronté:

- la question ponctuelle : appelle une réponse unique et immédiate : date, adresse, nom d'auteur, etc.
- la question chronologique : implique un déroulement dans le temps et donc un caractère rétrospectif
- la question de synthèse qui demande au documentaliste de rassembler des références permettant à l'utilisateur de faire la synthèse des connaissances acquises au jour de la demande
- la question « état de l'art » où ne seront sélectionnés que les documents d'actualité sur des recherches ou projets en cours.

Les vraies questions à se poser :

- quel est le thème de la recherche
- quel est le contexte
- pourquoi est-ce que je recherche ces informations

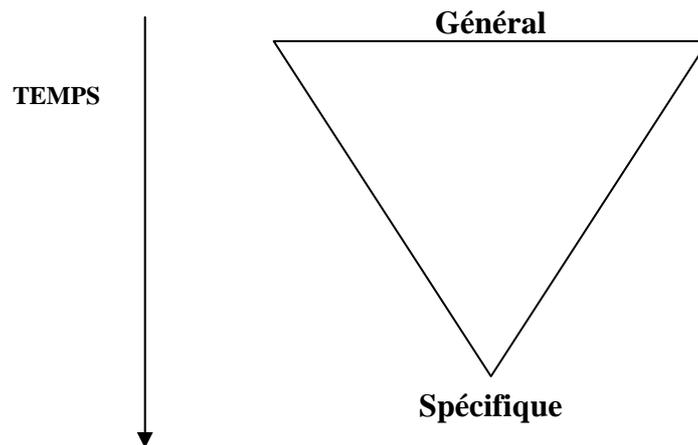
⁸⁶ Ressources pour l'autoformation à la recherche d'information :
<http://www.uhb.fr/urfist/Supports/FormaDoctor/FormDoctMethodo.htm#RessourcesAutoformation>
HEC, « méthode de recherche sur Internet », Laurence Loggh-Didier - Recherche d'information
http://www.accart.nom.fr/Cours/RechInfo_01.html - Méthodologie de la recherche documentaire
<http://users.l1vm-serv.net/aeris/cours/chercher/methodes.html>

- y'a t-il des synonymes à mes termes de recherche ?
- quel est le périmètre de ma recherche (langue, pays, date...) ?
- est-ce que je connais déjà des ressources existantes (Web et hors Web) sur le sujet (docs, contacts, etc.)?

VI.5.c Étapes clés de la recherche d'information :

1. **Remonter à la première expression du besoin en information** : là où se trouve la vraie question. En quelles circonstances s'est posé le problème : est-ce au cours d'une réunion professionnelle, d'une discussion entre collègues, en lisant un article paru dans la presse, ...
2. **Déterminer les "frontières" du problème à résoudre et analyser les diverses contraintes qui interviennent** ; les frontières peuvent être : la langue, l'origine des informations ou une période de temps considérée. Ce qui induit des contraintes, notamment dans l'appréhension du problème et dans le choix des outils.
3. **Structurer le problème, le décomposer en sous-problèmes et faire le point sur son propre savoir et sur son ignorance** (l'ignorance peut aussi être la méconnaissance d'une langue et il faudra, si besoin, proposer des solutions de traduction). C'est durant cette étape qu'il faut **dégager les concepts**, les **mots-clés** de la recherche et **élargir au maximum, dans un premier temps, le champ de sa recherche** pour ensuite circonscrire un à un les différents problèmes posés. **Il faut avoir en tête l'image d'un triangle renversé.**
4. **La quatrième étape est très importante** : il s'agit, arrivé à ce point de la réflexion, de **déterminer une stratégie et des tactiques de résolution du problème**. Il faut donc définir des priorités et dégager des critères de choix et d'évaluation.

Le visage de la recherche d'information :



VI.5.d Etablir une stratégie de recherche :

On peut partir d'un brainstorming : écrire tous les mots que suggère le sujet de la recherche.

Puis **utiliser les opérateurs logiques** (ET, OU, SAUF) [cf. partie « mode de recherche booléen ». Ces opérateurs offrent la possibilité d'exclure des caractéristiques, et d'en grouper d'autres pour une recherche, permet d'affiner la sélection sur une banque de données très rapidement. Les opérateurs logiques aident à la construction d'une définition exacte de la question. **Utiliser les troncatures** (*,\$,+ , ?), et les **opérateurs numériques** (: =,<,>)

VI.5.e Les conseils pour une meilleure recherche d'information sur le Web

Les étapes précédentes permettent d'être précis dans sa recherche. Il faut ensuite faire des choix: commencer par identifier la nature de la recherche: la recherche est-elle orientée "intention d'achat" ou est-elle purement informationnelle? Certains moteurs tels que Yahoo ! Mindset permettent à l'internaute de préciser son « intention de recherche », et le classement des résultats qui s'ensuit diffère selon une intention commerciale ou informationnelle.

Il faut ensuite multiplier les ressources, et ne pas utiliser un seul moteur de recherche, ce qui consisterait à regarder le monde au travers d'une seule fenêtre. Il faut multiplier les points de vue, même s'il peut être pratique d'utiliser un moteur de recherche favori de temps en temps, notamment lorsqu'on ne cherche pas une information trop importante. En revanche, dans le cadre d'une recherche d'information sérieuse comme pour une question de santé ou une question financière, il est bon de confronter les résultats obtenus à partir de différents moteurs avant de prendre une décision. Comparer les résultats et prendre des notes des résultats ressortant systématiquement en tête peut être une indication pour démasquer les liens commerciaux.

A ce sujet, si les métamoteurs offrent l'avantage de rendre plus rapide la recherche sur différents moteurs ils impliquent comme nous l'avons vu un biais dans l'affichage des résultats et il est important de garder cela à l'esprit lorsqu'on les utilise.

Ensuite, il ne faut pas se contenter de lire les 2 premières pages de résultats. Il faut lire davantage. L'information convoitée peut se trouver sur la douzième page, simplement parce que c'est une petite entreprise, ou un particulier qui la détient, et qu'il ou elle n'avait pas les moyens d'acheter une position en tête de liste.

Il faut être attentif aux indices et indications signifiant qu'un lien est commercial, des appellations telles que « liens sponsorisés », « site sponsorisé », « sponsored Listing », « sponsored Match », « sponsored Site », « sponsored Link » ou encore « site partenaire ». Au contraire, il faut rechercher les étiquettes « résultats Web » « Web results » ou « Web pages results » qui font généralement référence aux résultats générés à partir des algorithmes.

Une bonne chose à faire, en particulier lorsque l'on débute en recherche d'information, est de cliquer sur les liens « concernant la recherche » « About Search » « Info » « Aide » « Help » « Au sujet de ». Ces pages donnent généralement des informations sur la manière dont les moteurs gèrent les résultats commerciaux, elles donnent aussi des astuces et conseils pour utiliser au mieux le moteur en question. Il ne faut pas non plus hésiter à utiliser la fonction de recherche avancée.

Enfin, il peut être utile de connaître le jargon marketing et de se renseigner sur les modèles d'affaires liés à la recherche d'information Web. De nombreux articles paraissent chaque jour sur le sujet, un site de référence en France est www.abondance.com, on peut aussi visiter le site www.SearchEngineWatch.com ou www.Webrankinfo.com ou www.secrets2moteurs.com

Pour les futures recherches, il est bon de prendre le réflexe du marque-page, c'est-à-dire d'apprendre à poser un « bookmark » à partir du navigateur Web sur les sites que l'on a trouvés intéressants lors d'une recherche afin d'y accéder rapidement lors d'une recherche ultérieure.

VI.6 Web, langues et information

On compte environ 6000 langues parlées actuellement dans le monde, dont 1 000 utilisées par une population très faible en nombre. On estime que près de la moitié de ces 6000 langues sont menacées parce qu'elles sont parlées uniquement par des adultes qui ne les apprennent plus à leurs enfants.

La mort des langues n'est pas un phénomène nouveau. Depuis au moins 5000 ans, les linguistes estiment qu'au moins 30 000 langues sont nées et disparues, généralement sans laisser de trace. Mais le nombre de langues actuellement parlées dans le monde diminue à un rythme inégalé, de telle sorte que 90 % des langues existantes disparaîtront vraisemblablement au cours du prochain siècle. Il n'y aurait donc qu'environ 600 langues qui seraient relativement durables dont l'anglais, qui se répand de plus en plus, et qui est en voie de devenir la « lingua franca » mondiale, notamment via Internet.

Internet a été créé aux Etats-Unis en tant que réseau mis en place dès 1969 par le Pentagone⁸⁷ avant de se développer dans les organismes gouvernementaux et les universités. Historiquement, donc, la langue officielle du Web est l'anglais.

Cependant, de par sa vocation internationale, Internet doit être multilingue.

En effet, Internet représente un outil puissant capable d'annihiler les frontières plutôt que d'en créer de nouvelles.

Internet relie les gens dans le monde entier. Par là même il suggère une langue commune pour communiquer, cette langue apparaissant, de droit, comme étant l'anglais, puisque c'est la langue dans laquelle fut créé le réseau.

Oui, mais Internet permet de relier des personnes qui ne parlent pas la même langue et parmi elles, des personnes qui ne parlent pas anglais. Nous pouvons donc en conclure que si, effectivement, le fait de disposer d'une langue commune est appréciable, cela ne remplace en aucun cas la nécessité de disposer de toutes les langues possibles.

Internet favorise donc, dans le même temps, l'usage d'une langue commune et le multilinguisme. L'intérêt croissant pour les langues et le besoin qu'on en a stimulent de par le monde la création de cours de langues à distance (e-learning) et d'instruments d'aide linguistique (traducteurs, dictionnaires...), et rendus disponibles rapidement et à bon prix, grâce au réseau.

Les critiques qui sont émises à l'égard de l'anglais doivent selon moi être tempérées, d'autant que l'amalgame entre l'anglais, les anglophones, les Anglais, et les Américains, est facile à faire, et malheureusement, est souvent fait.

Plusieurs raisons appellent à la réserve concernant les reproches fait à l'anglais. Premièrement la domination de l'anglais sur les autres langues du monde n'est pas aussi forte que l'on croit. D'autre part, comme le dit Claude de Loupy, « De tout temps, des langues se sont imposées comme langues "universelles". Ainsi en a-t-il été du sumérien, de l'akkadien, du grec, du latin et du français (pour le monde occidental ou proche-oriental) [et] les autres langues n'ont pas disparu pour autant. »

L'argument disant que l'anglais est une langue commerciale, est lui aussi à tempérer, en effet, il est toujours plus simple de faire des affaires dans sa langue maternelle.

⁸⁷ sous le nom d'Arpanet (Advanced Research Project Agency) + Net (réseau)

La langue officielle du Web devrait devenir petit à petit une langue faite d'emprunts aux autres langues, au fil du temps, combinant des mots de français, d'anglais, de japonais, de chinois, etc., utilisés fréquemment et recouvrant des réalités communes à tous ces locuteurs.

Ce qui est réellement en jeu dans la « guerre » menée contre la langue anglaise relève plutôt de la peur d'être dépossédé de quelque chose.

Pour comprendre ce sentiment et pour identifier ce dont pourrait se voir dépossédées les nations non-anglophones, il faut comprendre ce que représente une langue pour une personne, pour une nation, pour des États.

VI.6.a Qu'est ce donc qu'une langue ?

Une langue est un système de signes linguistiques vocaux, graphiques ou gestuels qui permet la communication entre les individus. Une définition linguistique de la langue précise que c'est un système de signes doublement articulés, c'est-à-dire que la construction du sens se fait à deux niveaux d'articulation. On trouve tout d'abord celui des entités signifiantes (morphèmes et lexèmes, ou monèmes) formant les énoncés puis celui des unités distinctives de sens (phonèmes) formant les unités signifiantes. Ces deux niveaux d'articulation déterminent les premiers niveaux de la description linguistique : phonologie, morphologie et syntaxe.

On distingue généralement la langue (système de signes) et le langage (faculté humaine mise en œuvre au moyen d'un tel système). On distingue également, depuis Ferdinand de Saussure, la langue et la parole (c'est-à-dire l'utilisation effective du système de la langue par les locuteurs).

On appelle langue maternelle d'une personne la ou les langue(s) que cette personne a apprise(s) dans son enfance au cours de son apprentissage du langage.

Si la grande majorité des langues parlées dans le monde sont des langues naturelles, qui se sont formées spontanément à partir d'états de langue antérieurs, il existe cependant aussi des langues artificielles ou langues construites, comme l'espéranto, le volapük, l'ido l'interlingua, le lojban, ou encore le klingon, qui ont été créées consciemment par des individus.

On appelle langue morte ou éteinte une langue qui n'est plus pratiquée oralement comme langue maternelle, mais qui peut être encore utilisée dans certains domaines (tels que la religion). La connaissance des langues mortes, en permettant l'étude des textes anciens, est utile notamment à la linguistique historique, ainsi qu'à l'histoire et à ses disciplines annexes. Les deux langues mortes les plus importantes de la culture occidentale sont le latin et le grec ancien. L'expression « langue morte » est contestée par beaucoup de ceux qui les pratiquent. On lui préfère alors langue ancienne. Il est possible de « ressusciter » des langues mortes, comme le montre l'exemple de l'hébreu.

Une langue est dite vivante lorsqu'elle est utilisée oralement par des personnes dont elle est la langue maternelle, ou par une communauté suffisamment nombreuse — et de façon suffisamment intensive — pour permettre une évolution spontanée de la langue (cas de l'espéranto). Une langue vivante est rarement un système uniforme et

rigide, elle varie généralement selon le lieu géographique (dialectes), le milieu social (sociolectes) et les individus (idiolectes) et, bien sûr, selon le temps, ce qui fait que, considérée à un moment donné, une langue est toujours en évolution et contient plusieurs états. Par exemple, le système phonologique des langues est en évolution constante, ce qu'étudie la phonétique historique.

Faire vivre une langue c'est préserver un patrimoine car ce sont les langues qui permettent d'exprimer notre vision du monde.

D'autre part les langues aident à construire la cohésion sociale de chaque nation : en permettant aux citoyens d'acquérir un « chant commun », une culture, une identité commune et une identité individuelle⁸⁸.

Comme le dit Renaud Donnedieu de Vabres « la langue est également le premier instrument d'insertion sociale et professionnelle, non seulement pour nos citoyens, mais aussi pour tous ceux que nous accueillons chez nous, ces migrants dont l'intégration dans nos sociétés se fera d'abord par le partage d'une langue commune.

C'est pourquoi les langues sont les vecteurs de l'éducation et de la culture.⁸⁹ »

La langue si elle porte sa culture est également l'instrument qui permet de s'ouvrir aux autres cultures.

« Connaître les langues, c'est rencontrer l'autre, le respecter, comprendre les peuples, dialoguer avec eux [...] voilà ce que nous apporte, en plus de la maîtrise de notre propre langue, l'apprentissage des autres langues dans leur diversité. »

Il ne s'agit donc pas de faire mener une fronde anti-anglophone, mais plutôt de prôner une diversité culturelle qui passe par l'expression des différences linguistiques et par la représentation des différentes nations via leur système sémiotique, sur le média Internet.

D'autant plus que la technologie hypertexte, et les navigateurs (qui permettent de définir des préférences linguistiques) se prêtent totalement à la déclinaison et à l'affichage des contenus dans toutes les langues⁹⁰.

VI.6.b Permettre au Web de devenir multilingue

Voici quelques-uns des éléments qui, permettront au Web multilingue de devenir une réalité :

- **La démocratisation de la technologie de l'information**

La technologie des ordinateurs a longtemps été le seul domaine d'une élite "technicienne", à l'aise à la fois dans des langages de programmation complexes et en anglais, la langue universelle des sciences et techniques. A l'origine, les ordinateurs n'ont jamais été conçus pour manier des systèmes d'écriture ne pouvant être traduits en ASCII. Il n'y avait pas de place pour autre chose que les 26 lettres de l'alphabet anglais dans un système de codage qui, à l'origine, ne pouvait même pas reconnaître

⁸⁸ C'est particulièrement le cas pour les communautés linguistiques fortement minoritaires pour qui ce problème est devenu une question vitale.

⁸⁹ Extrait d'un discours de Renaud Donnedieu de Vabres donné à l'occasion de la réception en l'honneur de la Fédération Européenne des institutions linguistiques nationales, lundi 8 novembre 2004

⁹⁰ Grâce également à Unicode

les accents aigus et les trémas, sans parler de systèmes non alphabétiques comme le chinois.

Mais la tradition a été bouleversée, et la technologie popularisée. Des interfaces graphiques tels que Windows et Macintosh ont accéléré le processus. La stratégie de marketing de Apple et de Microsoft a consisté à présenter son système d'exploitation comme facile à utiliser par le client moyen. A l'heure actuelle cette facilité d'utilisation s'est étendue au-delà du PC vers le réseau Internet, si bien que maintenant même ceux qui ne sont pas programmeurs peuvent insérer des applets Java dans leurs pages Web sans comprendre une seule ligne de programmation.

- **La compétition des grandes sociétés pour une part de "marché global"**

L'extension de cette popularisation locale est l'exportation de la technologie de l'information dans le monde entier. La popularisation est maintenant effective à l'échelon mondial, et l'anglais n'est plus nécessairement la langue obligée de l'utilisateur. Il n'y a plus vraiment de langue indispensable, mais seulement les langues personnelles des utilisateurs. Une chose est certaine : il n'est plus nécessaire de comprendre l'anglais pour utiliser un ordinateur, de même qu'il n'est plus nécessaire d'avoir un diplôme d'informatique.

La demande des utilisateurs non anglophones et l'effort entrepris par les sociétés high-tech se faisant concurrence pour obtenir les marchés mondiaux ont fait de la localisation un secteur en expansion rapide dans le développement des logiciels et du matériel. Le premier pas a été le passage de l'ASCII à l'ASCII étendu. Ceci signifie que les ordinateurs commençaient à reconnaître les accents et les symboles utilisés dans les variantes de l'alphabet anglais, symboles qui appartenaient le plus souvent aux langues européennes. Cependant une page ne pouvait être affichée qu'en une seule langue à la fois.

- **Innovation technologique**

L'innovation la plus récente est UNICODE. Bien qu'il soit encore en train d'évoluer et qu'il ait tout juste été incorporé dans les derniers logiciels, ce nouveau système de codage traduit chaque caractère en 16 octets. Alors que l'ASCII étendu à 8 octets pouvait prendre en compte un maximum de 256 caractères, UNICODE peut prendre en compte plus de 65.000 caractères uniques et il a donc la possibilité de traiter informatiquement tous les systèmes d'écriture du monde.

Les instruments sont maintenant plus ou moins en place. Ils ne sont pas encore parfaits, mais on peut désormais surfer sur le Web en utilisant le chinois, le japonais, le coréen, et de nombreuses autres langues qui n'utilisent pas l'alphabet occidental. Comme Internet s'étend à des parties du monde où l'anglais est très peu utilisé, par exemple la Chine, il est naturel que ce soit le chinois et non l'anglais qui soit utilisé. La majorité des usagers en Chine n'a pas d'autre choix que sa langue maternelle.

- **La démocratie linguistique**

Dans un rapport de l'UNESCO du début des années 50, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur Internet dans sa langue maternelle pourrait bien être son équivalent à l'Age de l'information⁹¹. Si Internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Le considérer comme la chasse gardée de ceux qui, par accident historique, nécessité pratique ou privilège politique, connaissent l'anglais, est injuste à l'égard de ceux qui ne connaissent pas cette langue.

- **Le commerce électronique**

Bien qu'un Web multilingue soit souhaitable sur le plan moral et éthique, un tel idéal ne suffit pas pour en faire une réalité dépassant les limites actuelles. De même que l'utilisateur non anglophone peut maintenant avoir accès à la technologie dans sa propre langue, l'impact du commerce électronique peut constituer une force majeure qui fasse du multilinguisme la voie la plus naturelle vers le cyberspace.

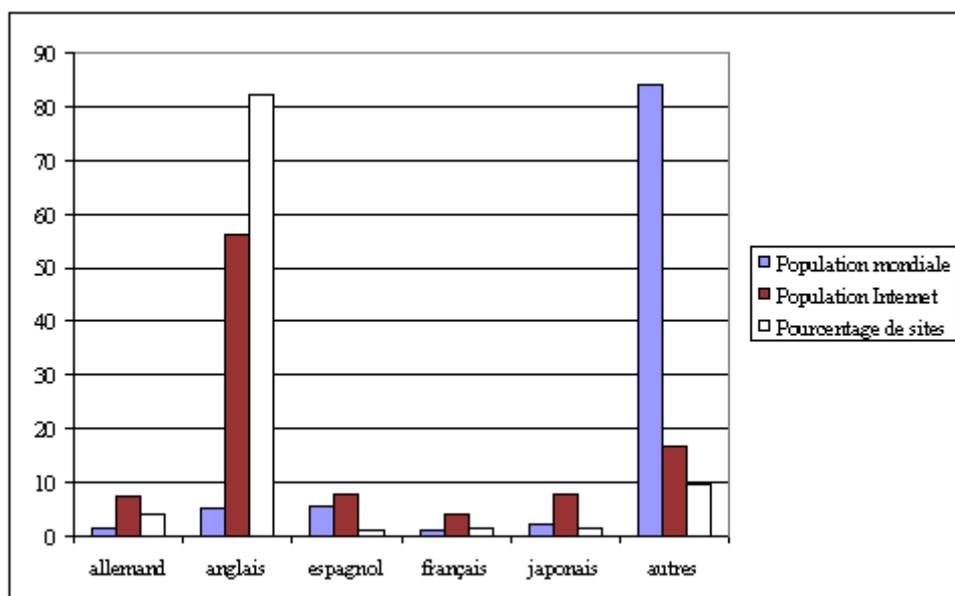
Les vendeurs de produits et services dans le marché virtuel mondial que devient Internet doivent être préparés à faire face à un monde virtuel qui soit aussi multilingue que le monde physique. S'ils veulent réussir, ils doivent s'assurer qu'ils parlent bien la langue de leurs clients, comme nous le disions plus haut.

⁹¹ L'Unesco a préparé un Projet de recommandation sur la promotion et l'usage du multilinguisme et l'accès universel au cyberspace. Elle a aussi mis en place l'Initiative Babel (1997) pour améliorer l'accès aux langues dans le cyberspace en utilisant les technologies de l'information et de la communication.

- **La répartition de la population linguistique sur Internet**

En 1997, Babel, une initiative conjointe d'Alis Technologies et de l'Internet Society, a mené la première étude d'ensemble sur la répartition réelle des langues sur Internet. Les résultats sont publiés dans le Palmarès des langues de la Toile, daté de juin 1997 et qui devrait être réactualisé prochainement. Les pourcentages étaient de 82,3 % pour l'anglais, 4 % pour l'allemand, 1,6 % pour le japonais, 1,5 % pour le français, 1,1 % pour l'espagnol, 1,1 % pour le suédois et 1 % pour l'italien.

Le schéma suivant indique la répartition, en pourcentage, de la population linguistique sur Internet et dans le monde (G.I.S., 1999⁹²) ainsi que la répartition linguistique des sites (Babel, 1997⁹³).



⁹² GIS (1999). *Global Internet Statistics (by language)*. [internet]. Cité par Claude De Loupy en septembre 1999 et disponible à l'époque sur <http://www.euromktg.com/globstats/>.

⁹³ Palmarès Babel des langues sur internet en 1997 disponible sur <http://alis.isoc.org/palmares.html>

Répartition linguistique de la population mondiale, Internet et des documents présents sur la Toile

Langue	Population mondiale	Population Internet	Pourcentage de sites
allemand	1,6	7,5	4
anglais	5,4	56,3	82,3
espagnol	5,6	7,7	1,1
français	1,2	3,9	1,5
japonais	2,1	7,8	1,6
autres	84,1	16,8	9,5

« On voit très nettement qu'en 1997, les proportions entre les communautés linguistiques n'étaient pas respectées lorsque l'on passait du monde réel à Internet. La prédominance des anglophones (56 %) sur la Toile était alors très nette, et les chiffres sont encore plus impressionnants en terme de documents disponibles (82 %) ». ⁹⁴

Le Top 10 des langues les plus parlées par les internautes⁹⁵

TOP 10 DES LANGUES SUR LE WEB	Nombre d'internautes, par langues	Pourcentage du nombre total d'internautes	Estimation de la population mondiale par langues	Taux de pénétration sur le Web
<u>Anglais</u>	310.743.832	31,9 %	1.107.807.851	28,1 %
<u>Chinois</u>	124.301.513	12,8 %	1.329.801.131	9,3 %
<u>Japonais</u>	78.050.000	8,0 %	128.137.485	60,9 %
<u>Espagnol</u>	62.959.431	6,5 %	389.587.559	16,2 %
<u>Allemand</u>	55.183.395	5,7 %	96.141.368	57,4 %
<u>Français</u>	40.124.900	4,1 %	374.555.140	10,7 %
<u>Coréen</u>	32.570.000	3,3 %	73.044.495	44,6 %
<u>Italien</u>	28.870.000	3,0 %	58.608.565	49,3 %
<u>Portugais</u>	28.792.000	3,0 %	227.628.673	12,6 %
<u>Danois</u>	14.655.328	1,5 %	24.685.673	59,5 %
LANGUES DU TOP 10	776.280.399	79,8 %	3.809.997.729	20,4 %
Reste des langues du monde	196.547.602	20,2 %	2.610.104.993	7,5 %
TOTAL MONDE	972.828.001	15,2 %	6.420.102.722	100,0 %

⁹⁴ Source : Solaris n°6 déc 1999 / jan 2000, De Loupy.

⁹⁵ Source : Internetworldstats, 21 novembre 2005

LES INTERNAUTES DANS LE MONDE ET STATISTIQUES DÉMOGRAPHIQUES⁹⁶

Regions du Monde	Population (2005 Est.)	Population % du Monde	Nombre d'internautes	Tx de croissance (internautes) 2000-2005	% Population (TxPénétration)	Internautes %Monde
Afrique	896.721.874	14,0 %	23.867.500	428,7 %	2,7 %	2,5 %
Asie	3.622.994.130	56,4 %	327.066.713	186,1 %	9,0 %	33,9 %
Europe	804.574.696	12,5 %	283.482.940	169,7 %	35,2 %	29,4 %
Moyen Orient	187.258.006	2,9 %	15.452.500	370,4 %	8,3 %	1,6 %
Amérique du Nord	328.387.059	5,1 %	223.971.489	107,2 %	68,2 %	23,2 %
Amérique Latine.Caraïbes	546.723.509	8,5 %	72.792.797	302,9 %	13,3 %	7,5 %
Australie	33.443.448	0,5 %	17.655.762	131,7 %	52,8 %	1,8 %
TOTAL MONDE	6.420.102.722	100,0 %	964.289.701	167,1 %	15,0 %	100,0 %

Le fait que l'anglais soit toujours représenté par un très fort pourcentage s'explique par les trois facteurs suivants : 1) l'anglais est la principale langue d'échange internationale, 2) les premières années ont vu la création d'un grand nombre de sites Web émanant des Etats-Unis, du Canada ou du Royaume-Uni, 3) la proportion de cybernautes est encore particulièrement forte en Amérique du Nord par rapport au reste du monde, pour les raisons suivantes : a) jusque très récemment, le matériel informatique et les logiciels étaient bien meilleur marché qu'ailleurs, b) on ne paie qu'un forfait mensuel pour les communications téléphoniques locales, ce qui rend l'utilisation d'Internet très économique par rapport à l'Europe, où les communications locales sont tarifées à la durée.

En contrepartie, même si l'anglais est encore prédominant, il faut cependant noter que de plus en plus de sociétés et d'organismes anglophones prennent conscience du fait que tout le monde ne comprend pas l'anglais. Les sites bilingues ou multilingues sont de plus en plus nombreux, ce pour des raisons aussi bien commerciales que culturelles

Sensible à ce problème, AltaVista, avait mis en place en décembre 1997 AltaVista Translation, un service de traduction automatisée de l'anglais vers cinq autres langues (allemand, espagnol, français, italien et portugais), et vice versa. Alimenté par des dictionnaires multilingues contenant plus de 2,5 millions de termes, ce service, gratuit et instantané, a été créé par Systran, société franco-américaine pionnière dans le domaine de la traduction automatique⁹⁷.

La traduction étant entièrement automatisée, elle est évidemment approximative. A défaut d'une « mémoire de traduction », ces outils permettent de prendre connaissance rapidement, et dans les grandes lignes, du contenu d'une page Web

⁹⁶ Source : Internetworldstats, 09 novembre 2005

⁹⁷ Le 9 mai 2000, Google a officiellement présenté une interface configurable en 10 langues, parmi lesquelles le français. Cette décision a été motivée par le fait que 35% des visiteurs sont des utilisateurs internationaux.

rédigée dans une langue que l'internaute ne connaît pas, mais sur laquelle l'outil dispose de ressources lexicales (lexiques, dictionnaires, etc.).

De nombreuses ressources sont disponibles sur le Web pour permettre aux internautes d'accéder à une information inscrite dans une langue autre que leur langue maternelle (cf. annexe⁹⁸).

Le site multilingue de Babel (allemand, anglais, espagnol, français, italien, portugais et suédois) comprend deux grands secteurs : langues (les langues du monde ; glossaire typographique et linguistique ; Francophonie), et Internet et multilinguisme (développer votre site Web multilingue ; le codage des écritures du monde).

Sur le site du W3C Internationalization/Localization (un secteur du World Wide Web Consortium (W3C), consortium international de l'industrie fondé en 1994 pour développer les protocoles communs du World Wide Web), on trouve une définition des protocoles utilisés pour l'internationalisation et la localisation : HTML (hypertext markup language), jeu de base de caractères, nouveaux attributs, HTTP (hypertext transfer protocol), négociation de la langue, URL (uniform resource locator) et autres identificateurs incluant des caractères non-ASCII (American standard code for information interchange). Le site propose aussi une aide pour créer un site multilingue.

- **L'importance de l'ingénierie linguistique**

Même si nous ne traiterons pas ici les difficultés techniques liées à la recherche d'information conceptuelles, d'idées, sur des chaînes de caractères, prenons tout de même le temps d'en caresser les contours.

Pour manipuler des mots sur le Web, les technologies doivent être capable d'extraire du sens à partir des chaînes de caractères traitées dans les contenus Web.

D'où l'importance de disposer d'outils capables de rechercher au-delà des apparences, en profondeur, rechercher des concepts.

Lancé en janvier 1999 par la Commission européenne, le site HLTCentral (HLT: Human Languages Technologies - Technologies des langues humaines) propose une courte définition de l'ingénierie du langage :

"L'ingénierie du langage permet de vivre en toute convivialité avec la technologie. Nous pouvons utiliser notre connaissance du langage pour développer des systèmes capables de reconnaître à la fois la parole et l'écrit, de comprendre un texte suffisamment en profondeur pour être capable de sélectionner des informations, de le traduire dans différentes langues et de générer aussi bien un discours oral qu'un texte imprimé.

L'application de ces technologies nous permet de repousser les limites actuelles de notre utilisation du langage. Les systèmes à commande vocale sont appelés à jouer un rôle prépondérant et à faire partie intégrante de notre vie quotidienne."

Marie Lebert⁹⁹ propose la définition suivante : « Le Génie Linguistique est une partie du traitement électronique de l'information. Dans ce contexte, le traitement

⁹⁸ Source : <http://www-Rocq.INRIA.Fr/qui/Philippe.Deschamp/divers/metalexis.html>

⁹⁹ MARIE LEBERT 1999, le multilinguisme sur le Web

automatique des langues est un secteur en plein développement. Il comprend les recherches et développement en matière d'analyse et de génération de textes, de reconnaissance, de compréhension et de synthèse de la parole. Il inclut les applications relatives à la gestion de documents, à la communication entre l'humain et la machine, à l'aide à la rédaction, à la traduction assistée par ordinateur. Il comporte des enjeux de type industriel et économique, de type scientifique et technologique mais présente aussi une dimension culturelle très spécifique. Il est toujours préférable de maîtriser la langue qui sert à exprimer les résultats des recherches, surtout si cette recherche porte sur la langue elle-même. »

La collaboration entre ingénieurs informaticiens et linguistes et la formation croissante de spécialistes se situant à l'intersection de disciplines aussi diverses que l'informatique, la linguistique, l'intelligence artificielle et la sémiotique, correspondent, et répondent, aux besoins en spécialistes du traitement de l'information qu'engendre la prolifération des contenus Web. Il leur appartient de faire du Web – entre autres choses – multilingue une réalité.

Les moteurs de recherche et le droit

Nous étudierons ici la réponse du droit face au détournement des modèles et aux litiges qui découlent de ces derniers. Dans ce modèle qui consiste à vendre des mots pour faire du commerce, comment sont considérées les marques, sont-elles des mots particuliers ? Quelles sont les conséquences des modèles d'affaires développés par les sociétés éditrices de moteurs de recherche sur les règles de la concurrence, quelles conséquences pour les consommateurs internautes ? Qui est tenu pour responsable des dérives et des abus ?

D'autre part, nous mettrons l'accent sur le fait que les informations circulant sur internet ne sont pas toutes libres de droit.

Quelles sont donc les difficultés que la législation peut rencontrer lorsqu'elle doit s'appliquer à un espace sans borne, et lorsque les différents continents ne fonctionnent pas tous selon le même régime juridique ?

Le réseau doit-il s'adapter au droit ou est-ce l'inverse ?

Nous reprendrons ici l'analyse de l'avocate Murielle Cahen qui apporte de nombreuses réponses sur ces problématiques.

VII Les moteurs de recherche et le droit

Le marché des outils de recherche n'a pas pu résister à l'effondrement du business modèle publicitaire basé sur la vente d'espaces et de bannières. La preuve en est que l'immense majorité d'entre eux, à la recherche de nouvelles sources de revenus, ont mis en place depuis près de huit ans de nouvelles offres de référencement : soumission payante sur les annuaires, référencement payant sur les moteurs, positionnement publicitaire sur les deux. En adoptant ce système, les engins de recherche sont passés d'une logique objective – pour les moteurs de recherche- ou éditoriale – pour les annuaires- à une technique purement commerciale, celle de la vente des mots-clés¹⁰⁰.

La méthode du référencement payant consiste, en effet, comme nous l'avons vu à offrir, moyennant finance, au site référencé une position en tête des résultats d'une recherche sur certains mots-clés prédéfinis. La rémunération se calcule soit par nombre de clic (Pay-Per-Clic), dans le cas du positionnement par enchères, soit par CPM (coût par milles pages vues), dans le cas du positionnement par achat du lien.

Le "Pay for Placement" et son corollaire, le nouveau marché de mots-clés, constituent, sans doute, des techniques de Web marketing fort puissantes, qui servent, notamment, des moyens pour accroître la visibilité des sites Internet et augmenter leur notoriété. Les entreprises, n'ont, naturellement, pas tardé à montrer leur intérêt. Selon une enquête, réalisé par le Journal du Net¹⁰¹, les deux-tiers des entreprises de plus de 500 personnes sont déjà au courant des nouvelles formules de référencement promotionnel, proposées par les outils de recherche majeurs (Overture, eSpotting ou Google) et plus de la moitié envisagent de faire leurs premiers pas sur ce support.

Face au succès que la technique du référencement promotionnel est censée avoir auprès des principaux acteurs de la Net économie, il convient d'examiner son niveau de conformité aux règles de droit existantes. Trois séries de questions se posent à cet égard :

Tout d'abord, l'éventuelle appropriation des mots-clés par les sites les plus riches et l'exclusion, sur certains engins de recherche, des sites disposant de ressources plus modestes, ne constituent-elles une entorse à la libre concurrence ?

Deuxièmement, la vente des mots clés par les moteurs de recherche est-elle une publicité clandestine ?

Enfin, comment peut-on protéger les titulaires des marques victimes de la pratique de " position squatting ", en d'autres termes, pourquoi les marques ne sont-elles pas des mots comme les autres ?

¹⁰⁰ http://www.murielle-cahen.com/p_positionnement.asp

¹⁰¹ www.Journaldunet.com : Liens promotionnels : près d'un tiers de cliqueurs réguliers

VII.1 Le positionnement payant face aux règles du droit de la concurrence.

Selon l'article 86 du Traité CE " Est incompatible avec le marché commun et interdit, dans la mesure où le commerce entre États membres est susceptible d'en être affecté, **le fait pour une ou plusieurs entreprises d'exploiter de façon abusive une position dominante sur le marché commun** ou dans une partie substantielle de celui-ci "

Le référencement payant et la vente des mots-clés sont susceptibles d'aboutir à des situations anti-concurrentielles, dans la mesure où l'appropriation des mots-clés par certaines entreprises peut empêcher les acteurs du même secteur d'être visibles sur le Net. Le dernier chapitre du rapport de l'OMPI¹⁰², relatif à la " gestion des noms et adresses de l'Internet : Questions de propriété intellectuelle ", évoque ce même problème :

" Plusieurs sociétés, sises dans plusieurs régions du monde, ont enregistré des noms de domaine comprenant tous l'élément " Télécom ". Ce sont Symmetricom, Inc (www.telecom.com), Telecom UK Ltd (www.telecom.co.uk), TWX Telecommunications gmbh (www.telecom.de), Telecom s.r.l. (www.telecom.it), Telstra Corporation Ltd. (www.telecom.com.au) et Swisscom (www.telecom.ch).

Néanmoins, certains des systèmes à mots clés existant actuellement ne renvoient l'utilisateur qui saisit le mot clé " telecom " qu'au site de Symmetricom, inc., sans rien dire des autres. (...)En saisissant le mot clé " golf ", l'utilisateur est dirigé d'une société sise aux États-Unis (www.golf.com) alors que d'autres utilisent le mot golf comme nom de domaine, en particulier une société du Royaume Uni (www.golf.co.uk), une société allemande (www.golf.de), une société néerlandaise (www.golf.nl), une société australienne (www.golf.com.au), etc. "

Dans sa décision du 9 juin 2000, le Conseil de la concurrence¹⁰³ a considéré, pour sa part, que lorsqu'il s'agit d'opérateurs exerçant leur activité sur le même marché, il est difficile de soutenir que chacun d'entre eux détient individuellement une position dominante, sauf s'il s'agissait d'une position dominante collective, ce qui n'était pas allégué. En l'espèce, n'étaient prouvés ni la position dominante collective ni l'abus, il n'en reste pas moins que ces deux conditions puissent être réunies à l'avenir.

En effet, la position dominante collective pourrait être établie avec la généralisation des accords conclus entre les prestataires de positionnement payant et les outils de recherche, de telle sorte que les résultats des premiers occupent les premières places de toute recherche effectuée. En France, par exemple, deux spécialistes en la matière, Ouverture et Espotting, ont déjà signé des accords avec, respectivement, AOL et Yahoo! Europe¹⁰⁴, par lesquels ces portails s'engagent à faire

¹⁰² " La gestion des noms et adresses de l'Internet : Questions de propriété intellectuelle ", Rapport final concernant le processus de consultations de l'OMPI sur les noms de domaine de l'Internet, 30 avril 1999, <http://wipo2.wipo.int>

¹⁰³ Décision n°2000-D-31 du Conseil de la concurrence, 9 juin 2000 relative à une saisine au fond et une demande de mesures conservatoires présentées par la société Concurrence, disponible sur Juriscom.net.

¹⁰⁴ Danny Sullivan, *Europe Paid Placement Warriors*, The Search Engine Rapport, 4 février 2002.

figurer les résultats fournis par les prestataires de positionnement payant partenaires en tête de liste.

Quant à l'abus de position dominante, le Conseil de la concurrence a rejeté l'argument de la société Concurrence en affirmant que " la fonction d'annuaire ou moteur de recherche sur Internet ne peut être tenue pour indispensable à la rencontre de la demande émanant du consommateur et l'offre de produits et services vendus sur Internet ". Ceci signifie, a contrario, que si l'apparition dans la liste des résultats des outils de recherche était nécessaire pour accéder au marché, en exclure un opérateur serait une entrave à la libre concurrence. Or, comme on l'a déjà montré dans une étude antérieure (cf. La responsabilité des outils de recherche), la croissance exponentielle du volume des données mises en ligne rend aujourd'hui nécessaire l'utilisation d'outils de recherche pour pouvoir accéder aux informations ou aux sites s'y trouvant. Sans eux l'Internet ne serait " qu'une sorte d'espace vierge au sein duquel il faudrait se déplacer sans carte ni boussole. Autrement dit, rien " ¹⁰⁵.

Certes, les outils de recherche peuvent choisir librement leur politique commerciale. La vente des mots clés ne paraît illicite face au droit de la concurrence, que si elle est réservée de manière exclusive à un *numerus clausus* d'opérateurs du marché électronique.

VII.2 Le positionnement payant : quelle protection pour les consommateurs ?

Les moteurs de recherche " traditionnels " affichent les résultats d'une requête selon un algorithme qui permet de classer les sites par ordre de pertinence. L'internaute présume, donc, comme nous l'avons vu dans le cadre de l'étude du ConsumerWebWatch, que les sites se trouvant en tête de liste sont ceux qui correspondent mieux à sa demande et c'est pour cette raison qu'il n'ira, très probablement, pas consulter la deuxième page. Lors d'un positionnement payant, se pose, alors, la question suivante : dans la mesure où la liste retournée à l'internaute suite à sa requête semble objective, n'est-il pas induit en erreur s'il ignore que la présentation des résultats n'est pas seulement gouvernée par la pertinence, mais aussi dirigée par le prestataire de référencement ?

Considérant qu'il s'agit là d'une question importante pour le commerce électronique, l'association américaine de consommateurs Commercial Alert a saisi l'année dernière la Federal Trade Commission (FTC), afin qu'elle se prononce sur la pratique de placement payant pratiquée par huit moteurs de recherche parmi les plus populaires (Altavista, HotBot, DirectHit, iWon, Looksmart, MSN.com, Netscape et Lycos). Selon l'association, ces derniers, en intégrant des liens commerciaux dans leurs listes sans les séparer des résultats " éditoriaux ", violent les règles fédérales qui bannissent la publicité clandestine.

¹⁰⁵ DIMEGLIO, Arnaud, *Les contrats de référencement dans l'Internet* ", Comm.Com.Elect., mars 2001, p.14. Tempérons toutefois cette affirmation en rappelant que l'on peut toujours accéder aux sites dont on connaît l'adresse par liens directs.

Après l'analyse des programmes de placement payant pratiqués par les outils en cause, la Federal Trade Commission a récemment donné sa position sur le phénomène.¹⁰⁶ Plus précisément, elle a considéré qu'il n'est pas opportun d'agir contre ces moteurs de recherche, vu notamment que certains d'entre eux distinguent clairement les résultats payants des autres. Néanmoins, dans sa lettre, outre qu'elle reconnaît que dans certains cas le consommateur peut être induit en erreur, la FTC souligne la nécessité d'indiquer clairement la nature du résultat communiqué, en retenant comme critère le fait que ceci n'aurait pas pu figurer dans la liste autrement que par le paiement.

VII.3 L'interdiction de la publicité clandestine ou trompeuse.

L'article 2 du décret du 27 mars 1992 relatif à la publicité et au parrainage audiovisuel prévoit que : " constitue une publicité toute forme de message télévisé diffusé contre rémunération ou autre contrepartie en vue de promouvoir la fourniture de biens ou de services ".

Par ailleurs, la directive " Commerce Electronique " du 8 juin 2000 énonce dans son article 2 que constitue une communication commerciale " toute forme de communication destinée à promouvoir, directement ou indirectement, des biens, des services, ou l'image d'une entreprise, d'une organisation ou d'une personne ayant une activité commerciale, industrielle, artisanale ou exerçant une profession réglementée "

Les moteurs de recherche sont des supports de publicité, puisqu'ils font la promotion des biens ou des services en échange d'une rémunération fournie par le site positionné. La jurisprudence, d'ailleurs, retient une conception large de la publicité : " le critère essentiel de la publicité réside dans le fait qu'il puisse véhiculer un message commercial, qu'elle qu'en soit la forme " ¹⁰⁷.

Le positionnement payant étant une publicité, il doit se conformer aux règles existantes en la matière. Le décret de 1992 précité dispose que " la publicité clandestine est interdite ". Par ailleurs, la loi de 1986 sur l'audiovisuel énonce dans son article 43 al.2 que " les messages publicitaires doivent être mentionnés comme tels ". L'article 6a de la Directive " Commerce Electronique ", enfin, prévoit que " la communication commerciale doit être clairement identifiable comme telle ".

Par conséquent, les engins de recherche doivent indiquer de manière claire la nature commerciale des liens figurant dans la liste des résultats, afin que l'internaute puisse savoir si un site a payé pour y figurer ou si sa présence sur la liste est le résultat d'une recherche réelle et objective.

Le Code de la consommation interdit de même " toute publicité comportant, sous quelque forme que ce soit, des allégations, indications ou présentations de nature à induire en erreur, lorsque celles-ci portent sur la nature des biens ou services qui font l'objet de la publicité ". Au délit de publicité clandestine s'ajoutera, donc, celui de

¹⁰⁶ Lettre de la Federal Trade Commission, Août 2002, disponible dans Juriscom.net

¹⁰⁷ CA Rennes, réf., 30 mars 2000, JCP ed.E 2000, n°48, p.1902

publicité trompeuse, lorsque l'annonceur a acheté un ou plusieurs mots clés qui ne correspondent pas à son activité commerciale réelle.

VII.4 L'application de la loi Sapin aux intermédiaires de positionnement payant.

La loi du 29 janvier 1993, dite loi Sapin¹⁰⁸ s'applique au support Internet, qu'il s'agisse d'espaces dans les bandeaux, icônes ou liens, dès lors qu'ils passent par l'intermédiaire des serveurs ou vendeurs d'espaces ou bien encore d'agences de publicité.

Le positionnement payant passe par l'intermédiaire d'un tiers, lorsque le responsable du site intéressé s'adresse non pas directement à l'outil de recherche, mais à un prestataire de référencement, qui lui se met en contact avec ce dernier. Ces prestataires de référencement apparaissent, alors, comme des intermédiaires dans l'achat d'espace publicitaire, ce qui implique l'application de la loi Sapin. Ainsi, un contrat de mandat devra être établi par écrit entre le prestataire de positionnement et le site annonceur. Il est encore prévu que l'intermédiaire ne pourra percevoir d'autres rémunérations que celles figurant au contrat de mandat, ce qui signifie que le prestataire ne peut être rémunéré que par l'annonceur et non pas par l'outil de recherche.

Il en va, pourtant, autrement, lorsqu'on est en présence d'un contrat de partenariat entre différents outils de recherche ou entre un outil de recherche et un site portail. Dans ce dernier cas, le contrat n'a pas pour objet l'achat d'espace publicitaire, mais la diffusion des résultats d'un outil de recherche par un autre outil partenaire ou par un site, en échange, bien sûr d'une rémunération. Ainsi, la loi Sapin ne semble pas devoir s'appliquer en l'espèce.

VII.5 Le " position squatting " et la protection des marques.

Le position squatting est " le fait de payer pour apparaître dans les premiers résultats, lors d'une recherche sur un mot clé représentant une marque, dont on détient pas les droits "¹⁰⁹. Selon une étude réalisée récemment par une société de référencement¹¹⁰, 60% des entreprises du CAC 40 en sont victimes.

Le titulaire d'une marque peut-il s'opposer à ce que d'autres personnes utilisent celle-ci en tant que mot-clé pour acheter une position privilégiée sur les résultats des recherches des internautes ? Quels sont les recours dont celui-ci dispose vis à vis tant du squatteur que de l'outil de recherche ?

¹⁰⁸ Loi n° 93122 du 29 janvier 1993 relative à la prévention de la corruption et à la transparence de la vie économique et des procédures publiques, JO du 30 janvier 1993.

¹⁰⁹ Laurent Bailliard, " *Quand le référencement flirte avec l'illégalité* ", Vnuned.fr 3 janvier 2002

¹¹⁰ " Le Position squatting, nouveau défi des marques sur Internet ", Cvfm.com, décembre 2001.

VII.5.a La responsabilité du squatteur

- **L'action en contrefaçon de marque.**

L'article L.713-2 du Code de la propriété intellectuelle interdit la reproduction, l'usage ou l'apposition d'une marque, sans l'autorisation préalable du titulaire de celle-ci, ainsi que "l'usage d'une marque reproduite, pour des produits ou services identiques à ceux désignés dans l'enregistrement".

L'achat, auprès d'un outil de recherche, d'une position sur un mot-clé constituant un signe protégé peut-il être considéré comme un usage de marque au sens de l'article 713-2 du CPI, même si, comme c'est le cas très souvent, le signe en cause n'est pas visiblement utilisé ?

La réponse semble être positive, compte tenu, notamment, de la jurisprudence relative à la protection des marques face à leur reproduction en tant que méta-tags dans le code source d'un site. Celle-ci considère, en effet, que l'insertion par un tiers d'un signe distinctif déposé à titre de marque dans les balises méta d'une page Web peut être qualifiée de contrefaçon¹¹¹. Cette solution a été confirmée récemment dans une affaire opposant deux sociétés concurrentes dans le commerce des orgues de barbarie. En l'espèce, le TGI de Paris¹¹² a considéré que "la reproduction sans autorisation par une société de la marque d'en de ses concurrents sur la page source de son site Internet était constitutive de contrefaçon et d'atteinte au nom commercial de la société concurrente".

Dans le cas du position squatting l'effet escompté est identique : faire apparaître son site dans les résultats des recherches sur les mots-clés choisis. Le fondement de la contrefaçon a, d'ailleurs, été utilisé de manière reconventionnelle dans une affaire opposant aux Etats-Unis deux sociétés titulaires de la marque Nissan (Nissan motor co. ltd et Nissan computer corporation)¹¹³, dont l'une se plaignait du fait que l'autre apparaissait avant elle dans les résultats des recherches sur le mot-clé "Nissan" et "Nissan.com". Le tribunal américain a, pourtant, rejeté la demande¹¹⁴.

Qu'en est-il, pourtant, si l'acheteur du mot-clé, protégé à titre de marque, est en droit d'utiliser cette dernière pour annoncer la vente des produits qu'il a acquis licitement ?

¹¹¹ Voir par exemple les affaires *SARL Distrimart c/ SA Safi*, CA Paris, 14 mars 2001, disponible sur *Juriscom.net*, <http://www.juriscom.net/txt/jurisfr/ndm/caparis20010314.htm> et *SA Citycom c. SA Chanel*, CA Paris, 3 mars 2000, disponible sur *ForumInternet.org*, <http://www.forumInternet.org/documents/jurisprudence/lire.phtml?id=79>

¹¹² TGI Paris, 29 octobre 2002, décision disponible sur le site www.legalis.net/

¹¹³ United States District Court, Central District of California, *Nissan Motor Co. Ltd v. Nissan Computer Corporation*, 29 août 2001

¹¹⁴ Pour une étude plus analytique de l'affaire voir <http://searchenginewatch.com/sereport/02/02-bodysolutions.html>

En effet, rien ne s'oppose a priori à ce qu'un vendeur membre d'un réseau de distribution utilise la marque comme élément de référencement de son site¹¹⁵. La jurisprudence reconnaît au distributeur des produits authentiques le droit d'utilisation de la marque pour la publicité de ses produits. La CJCE a été amené à la même conclusion par le biais de la règle de l'épuisement des droits¹¹⁶. Dans l'affaire opposant la société BMW à un garagiste indépendant, la CJCE a, par ailleurs, jugé que le titulaire d'une marque ne pouvait "interdire à un tiers l'usage de sa marque en vue d'annoncer au public qu'il effectue la réparation et l'entretien des produits de cette marque..." .Selon cette décision, la liberté du revendeur d'utiliser la marque dans la promotion a une limite, lorsque la marque est utilisée d'une "manière telle qu'elle peut donner l'impression qu'il existe un lien commercial entre le revendeur et le titulaire de la marque et notamment l'entreprise du revendeur appartient au réseau de distribution et qu'il existe une relation spéciale entre eux".

En revanche, dans l'affaire Citycom c. Chanel¹¹⁷, la Cour d'appel de Paris a condamné pour contrefaçon de marque un distributeur parallèle qui revendait des produits Chanel en ligne, au motif qu'il avait utilisé la marque dans le code source de son site. Elle a souligné que "ce code permettait aux internautes, par le biais des annuaires et des moteurs de recherche d'accéder directement au site susceptible de les intéresser par la simple opposition de la marque comme mot-clé".

Le titulaire de la marque victime de position squatting pourra, donc, intenter une action pénale en contrefaçon et se porter partie civile, devant le tribunal correctionnel. Il peut également intenter une action en référé spécifique au droit des marques sur la base de l'article L.716-6 du Code de la propriété intellectuelle, afin que le juge ordonne au contrefacteur de ne plus utiliser la marque en tant que mot-clé pour occuper une position privilégiée dans les résultats des outils de recherche.

¹¹⁵ Quant à la question de savoir si un distributeur agréé peut vendre les produits du réseau en ligne, il faut tout d'abord examiner les stipulations relatives du contrat de distribution. Le paragraphe 51 des lignes directives pour la mise en œuvre du Règlement 25/90/99 est très net pour ne pas exclure a priori la vente en ligne dans le cas des ventes de distribution. Le fournisseur peut imposer des normes de qualité pour la présentation du site, les promotions etc., mais le texte souligne que l'interdiction de vente sur Internet n'est licite que si elle objectivement justifiée. La Cour d'appel de Versailles, dans son arrêt *Fabre c/ Alain B.*, n'a pas exclu qu' "à l'avenir, ce nouveau mode de distribution puisse s'intégrer dans un réseau de distribution sélective, avec des critères de qualité à définir". Décision disponible sur Juriscom.net, <http://www.juriscom.net/txt/jurisfr/ce/caversailles19991202/htm>

¹¹⁶ CJCE, 4 novembre 1997, *Dior c. Evora*, PIBD 1999 n°676, III, p.221.

¹¹⁷ CA Paris, 3 mars 2000, disponible sur ForumInternet.org <http://www.forumInternet.org/documents/jurisprudence/lire.phtml?id=79>

- **L'action en concurrence déloyale.**

Selon le principe de spécialité, le titulaire de la marque ne peut empêcher l'usage de celle-ci pour désigner des produits et services différents de ceux visés dans l'enregistrement. Ainsi, lorsque le site positionné n'est pas concurrent du titulaire de la marque, ce dernier ne peut pas tenter une action en contrefaçon de marque.

Cependant, le titulaire de la marque victime de position squatting peut toujours tenter une action en concurrence déloyale contre le squatteur, si celui-ci ne cherche qu'à " profiter du travail d'autrui sans bourse délier " ¹¹⁸ et notamment lorsque les mots-clés réservés n'ont aucun rapport avec l'activité commerciale de ce dernier. Pour cela il faut établir qu'il y a une faute, un dommage et un lien de causalité entre les deux. La faute résultant du comportement déloyal, le préjudice consistera en la perte de clientèle, l'atteinte à l'image ou à la valeur de la marque.

Il en va autrement, lorsque la marque squattée est notoire et que le squatteur cherche à profiter du renommé de celle-ci pour augmenter sa visibilité. Selon l'article 713-5 du Code de la propriété intellectuelle, " L'emploi d'une marque jouissant d'une renommée pour des produits ou services non similaires à ceux désignés dans l'enregistrement engage la responsabilité civile de son auteur s'il est de nature à porter préjudice au propriétaire de la marque ou si cet emploi constitue une exploitation injustifiée de cette dernière " .

¹¹⁸ CA Paris, 22 février 1995, D.S., 1996, Somm.250.

VII.5.b La responsabilité de l'outil de recherche.

L'outil de recherche qui vend des mots-clés peut-il voir sa responsabilité engagée au même titre que le squatteur ?

La responsabilité des fournisseurs des outils de recherche n'est traitée ni dans la Directive Commerce Electronique 2000/31/CE ni dans la loi du 1 août 2000 concernant les intermédiaires de l'Internet. Toutefois, le législateur européen a chargé la Commission de présenter un rapport, avant le 17 juillet 2003, sur la nécessité de présenter des propositions relatives à la responsabilité des " services de moteurs de recherche ".

La directive " commerce électronique " précitée, ne retient la responsabilité du prestataire d'hébergement que s'il est prouvé que celui-ci a eu connaissance de l'activité illicite et n'a rien entrepris pour la faire cesser¹¹⁹. Dans une étude antérieure, on a proposé de faire le même raisonnement pour les outils de recherche. En effet, étant donné le rôle joué par ceux-ci dans la société de l'information, leur responsabilité ne peut être plus étendue que celle des fournisseurs d'accès et d'hébergement. Ainsi, le TGI de Paris¹²⁰ n'a pas retenu la responsabilité du moteur de recherche AltaVista, car il avait retiré le site qui portait atteinte aux droits de la personnalité de M. Delanoë dès qu'il en avait eu connaissance.

Doit-on mener le même raisonnement à propos des outils qui propose un service de positionnement payant ? La réponse semble être plutôt négative, compte tenu du fait que, dans ce cas précis, l'outil de recherche se présente non pas comme un prestataire technique, mais en tant que support de publicité qui vend son espace.

En droit civil, le fournisseur du positionnement payant peut être tenu responsable, sur le fondement de l'article 1382 du Code civil, si en connaissance de cause il a vendu des mots-clés déposés à titre de marques.

En droit pénal, l'outil de recherche pourrait voir sa responsabilité engagée, soit comme auteur direct d'une infraction, soit, plus probablement, comme complice, sur le fondement de l'article 121-7 du Code pénal, pour avoir fourni en connaissance de cause une aide à la commission de l'infraction.

¹¹⁹ Art. 15 de la directive.

¹²⁰ TGI Paris, ref., 31 juillet 2000, *M. Bernard D. c/ Sté AltaVista Company*, disponible sur Juriscom.net : www.juriscom.net/txt/jurisfr/cti/tgiparis20000731.htm

VII.5.c La responsabilité civile délictuelle du prestataire de positionnement payant.

Le droit commun de la responsabilité est fondé sur la notion de faute au sens de l'article 1382 du Code civil. Ainsi, le fournisseur d'un outil de recherche peut être tenu responsable civilement des dommages causés, du fait de la vente d'un mot-clé, à une tierce personne. Dans ce cas, cette dernière devra prouver la réalité du dommage subi, la faute du fournisseur de l'outil de recherche et un lien de causalité entre les deux.

Or, la faute de l'outil de recherche sera, le plus souvent, difficile à démontrer. Certes, dans le cas où la marque serait notoire, la responsabilité de l'outil pourrait être engagée en raison de sa connaissance du signe squatté. Quid, pourtant, s'il s'agit d'une marque non connue ?

La jurisprudence, quant à elle, se montre hésitante. En effet, si le moteur de recherche Excite a été condamné en Allemagne pour avoir vendu à la société Fragrance Counter des mots-clés déposés à titre de marques par la société Estée Lauder¹²¹, sa responsabilité n'a pas été retenue par le tribunal américain, dans l'affaire Playboy, au motif que " Playboy " était devenu un terme générique sur Internet. Actuellement est en cours aux Etats-Unis une nouvelle affaire, opposant la société Mark Nutritional Incorporation, titulaire de la marque Body Solutions à quatre outils de recherche (AltaVista, FindWhat, Kanoodle et Overture), qui, eux, risquent de voir leur responsabilité engagée pour avoir vendu la marque en cause à des concurrents de la société demanderesse¹²².

Certains auteurs¹²³ envisagent, même, la possibilité d'engager la responsabilité pour risque des moteurs de recherche, afin de permettre au titulaire d'une marque victime de position squatting de réparer le préjudice subi du fait de l'utilisation de celle-ci, à titre de mot-clé, dans les moteurs de recherche par une tierce personne.

¹²¹ Landericht Hamburg, 16 février 2000, Estée Lauder c. Fragrance Counter

¹²² Voir <http://searchenginewatch.com/sereport/02/02-bodyolutions.html>

¹²³ André Lucas, Jean Deveze et Jean Frayssinet, *Droit de l'informatique et de l'Internet*, PUF, novembre 2001, n° 704.

VII.5.d La responsabilité pénale du prestataire de positionnement payant.

En matière pénale, l'outil de recherche, s'il ne peut pas être considéré comme auteur de contrefaçon de marque, il peut être condamné pour complicité, s'il a facilité la commission du délit¹²⁴, en vendant, en connaissance de cause, à un tiers non-titulaire un mot-clé protégé par le droit des marques. Il en va de même s'il a été prévenu par le titulaire de la marque de la contrefaçon et, malgré cela, il n'a pas enlevé le positionnement en cause.

En conclusion, on s'aperçoit que la technique du positionnement payant et de la vente des mots clés est loin d'être à l'abri des procédures judiciaires : abus de position dominante de la part des outils de recherche, publicité clandestine et/ou trompeuse, contrefaçon de marque et parasitisme sont des allégations qui, une fois invoquées, constituent une véritable menace pour les engins de recherche. De plus, tenter une action contre ceux derniers pour acquérir réparation du préjudice subi présente un avantage, lorsque l'acheteur du positionnement se prouve insolvable. Tout cela rend nécessaire d'effectuer un contrôle a priori sur les mots-clés en vente et sur la qualité des futurs acheteurs, afin d'assurer que ceux-ci sont en droit d'utiliser la marque qui se cache souvent derrière les mots-clés choisis.

¹²⁴ Art. 121-7 al.1 Nouveau Code Pénal

VII.6 Le droit d'auteur¹²⁵

- **Les œuvres protégées sur Internet**

En vertu de l'article L112-1 du Code de la propriété intellectuelle, les droits d'auteurs s'appliquent à toute œuvre de l'esprit, quels qu'en soient le genre, la forme d'expression, le mérite ou la destination. En conséquence, toutes les données ou informations que l'on rencontre sur Internet ne sont pas protégées par le droit d'auteur, mais le champ d'application de la propriété littéraire et artistique demeure très large, car il comprend toutes les oeuvres de l'esprit à caractère original, c'est-à-dire empreintes de la personnalité de leur auteur.

Ainsi, les textes, de toute nature, diffusés sur le réseau (extraits d'ouvrages littéraires ou scientifiques, articles journalistiques, discours publics...) sont protégés par le droit d'auteur. De même pour les images fixes ou animées (photographies, reproductions d'oeuvres d'art, images de synthèse...), pour la musique (mais les sons en tant que tels ne sont pas protégés), et pour toute œuvre audiovisuelle, c'est-à-dire toute œuvre constituée d'une séquence animée d'images sonorisées ou non.

Le Multimédia est une catégorie d'œuvre, récemment apparue, qui bénéficie du droit d'auteur et dont une des définitions se trouve dans le rapport Théry de 1994 sur les autoroutes de l'information: « Le multimédia est un ensemble de services interactifs utilisant le seul support numérique, pour le traitement et la transmission de l'information dans toutes ses formes : textes, données, sons, images. ». Ainsi la catégorie d'œuvre multimédia intègre les CD-ROMS interactifs, mais aussi les sites Web sur Internet. Généralement, l'œuvre multimédia est une œuvre composite, au sens du Code, c'est-à-dire une œuvre originale dans laquelle une œuvre préexistante a été incorporée sans la collaboration de l'auteur de cette dernière. Le régime juridique de l'œuvre composite établit qu'elle est la propriété de l'auteur qui l'a réalisée, sous réserve des droits de l'auteur de l'œuvre préexistante.

Les logiciels sont entrés dans la catégorie des œuvres protégées par le droit d'auteur grâce à la loi du 3 juillet 1985. De plus, en 1994, le législateur a transposé la directive européenne du 14 mai 1991 relative à la protection juridique des programmes d'ordinateurs, et qui permet de protéger non seulement le logiciel, mais également les travaux préparatoires de conception aboutissant au développement du programme. De plus, afin de lutter contre le pillage des oeuvres informatiques, le Code dispose que toute reproduction autre que la copie de sauvegarde établie par l'utilisateur, ainsi que toute utilisation d'un logiciel sans autorisation expresse de l'auteur est rigoureusement illicite.

¹²⁵ http://memoireonline.free.fr/propriete_intellectuelle.htm

VII.6.a Difficultés d'application des règles juridiques sur Internet

- **Difficultés dans l'application du droit d'auteur face à certaines catégories d'œuvres**

Les droits d'auteur apparaissent parfois comme inadaptés aux nouvelles catégories d'œuvres ou simplement à la numérisation de catégories plus anciennes. Sans nécessiter la création d'un droit spécifique pour ces catégories, il convient de relever les aménagements réglementaires qui ont été réalisés ou devraient l'être, sous peine de voir la loi inapplicable ou bien de paralyser l'activité créatrice des plus entrepreneurs.

- **Les œuvres multimédia**

Un produit multimédia apparaît comme un produit d'une nouvelle génération, qui doit être protégé, bien qu'il ait du mal à s'intégrer dans les schémas classiques du droit d'auteur.

- L'imprécision dans la cession des droits d'auteur

Le produit multimédia étant généralement le fruit du travail d'un ou plusieurs salariés, il faut s'interroger sur la protection des auteurs liés par un contrat de travail, et en particulier sur les conditions du transfert des droits patrimoniaux à l'employeur. Pendant longtemps, une partie de la jurisprudence a adopté une position favorable à l'employeur, malgré les dispositions légales, en admettant l'idée d'une cession implicite dès lors que celle-ci est limitée à l'activité normale de l'entreprise, et que les salariés soient auteurs d'œuvres créées dans le cadre d'un contrat de travail. La Chambre civile de la Cour de Cassation semble avoir mis fin à cette position dans un arrêt du 16 Décembre 1992. Il faut donc considérer qu'à priori, il ne peut y avoir ni cession automatique, ni cession tacite des droits du salarié. Une clause particulière du contrat de travail devra éclaircir cette situation.

- Les difficultés dans la détermination des auteurs

En matière de jeu sur support multimédia, la difficulté réside dans la détermination, parmi tous les intervenants sur le jeu, de ceux qui peuvent revendiquer la qualité d'auteur. Par exemple, l'auteur du logiciel, c'est-à-dire du moteur informatique ne peut revendiquer le statut d'auteur. Par contre, pour les auteurs à l'origine de la création du jeu multimédia, leur qualité d'auteur découlera du type de prestation qu'ils auront effectué, ce qui est assez imprécis et difficile à évaluer. Ainsi, un infographiste peut réclamer le statut d'auteur, si son intervention ne se borne pas uniquement à de la retouche d'images, mais qu'il élabore de toutes pièces des images et des séquences animées. L'éditeur de jeu doit donc, dès l'élaboration du jeu, définir par contrat la mission de chacun. Cependant, il doit éviter de conférer le statut d'auteur à trop d'intervenants au risque de voir se multiplier les sources de litiges, notamment en matière de droit moral.

- **Les logiciels**

Dans leur grande majorité, les logiciels et programmes d'ordinateurs sont aujourd'hui des créations réalisées par des salariés. La loi a ici attribué les prérogatives patrimoniales du droit d'auteur à l'employeur et gelé les prérogatives morales de l'auteur, se démarquant du droit d'auteur classique pour se rapprocher de la notion de copyright.

En effet, par la loi 10 mai 1994 du code de la Propriété Intellectuelle, l'employeur est le titulaire du logiciel créé par un employé dans l'exercice habituel de son activité professionnelle ou à la suite de recherches spécifiquement confiées à l'employé et qui n'entrent pas dans ses fonctions habituelles. Le salarié reste certes investi de son droit moral sur sa création, mais celui-ci se limite à la faculté de revendiquer la paternité de la conception et de la réalisation.

D'autre part, les logiciels sont théoriquement protégés par le droit d'auteur et non par le droit des brevets mais, dans la pratique, l'Office Européen des Brevets accepte parfois de breveter une invention utilisant un logiciel et l'accord APDIC (Accord sur les aspects des droits de propriété intellectuelle qui touchent au commerce) n'exclut pas non plus la brevetabilité des logiciels. La situation actuelle accepte donc les deux types de protection, mais il semble nécessaire de clarifier cette situation, qui peut être à l'origine d'ambiguïtés notamment au niveau du cumul des deux protections.

- **Les bases de données**

Si la mise en place d'une base de données est un projet lourd financièrement, et important pour la productivité de l'entreprise, on comprend alors la nécessité d'une protection juridique efficace et adaptée à ce type d'investissement.

Le Parlement et le Conseil européen ont adopté le 11 mars 1996 une directive élargissant le droit d'auteur aux bases de données et fondée sur un régime particulier, qui dissocie le contenu et la structure d'une telle œuvre. Le contenu d'une base de données, ne présentant aucune originalité (puisque constitué de données chiffrées ou factuelles), est protégé par un droit sui generis ; la structure, se concevant d'avantage comme le fruit d'un travail intellectuel original, est protégée par le droit d'auteur. En effet, la sélection des informations et l'organisation du contenu correspondent à l'expression d'une logique, propre à l'auteur de la base. A titre d'illustration, la Cour d'appel de Paris a considéré, à propos d'un annuaire, qu'il était protégeable non pour les adresses qui le composent, mais du fait de la « présentation générale qui en est faite ».

L'esprit de ce droit spécifique consiste donc en la protection des données en tant que source d'informations, et ce mécanisme à pour vocation de pallier à l'inadéquation du droit d'auteur dans la défense des investissements économiques d'un créateur de base de données

- **Les photographies sur Internet**

La mise en ligne d'une photographie met en jeu à la fois le droit patrimonial de l'auteur et son droit moral. La consultation de sites Web fait apparaître que ce dernier est souvent malmené : le nom de l'auteur est souvent oublié, les photos sont modifiées ou recadrées sans autorisation, cela en raison des nombreuses possibilités offertes par l'ordinateur.

Une adaptation du droit moral dans le secteur, en pleine expansion, de l'imagerie semble donc nécessaire, au risque de voir la règle juridique violée et ridiculisée.

Ainsi la réduction d'une photographie induite par sa numérisation et pour des raisons techniques (de manque de place), ne devrait pas constituer une atteinte au droit moral de l'auteur. De même pour le recadrage d'une photo effectué pour des raisons techniques.

En ce qui concerne le pillage du droit patrimonial de l'auteur d'une photographie, un moyen simple peut consister, lors de la première divulgation sur le réseau, en l'utilisation d'une faible résolution de l'image, afin que sa réutilisation soit dépourvue de tout intérêt esthétique. Ce mécanisme purement informatique autoriserait les auteurs d'oeuvres visuelles à mettre en ligne un certain nombre de leurs créations, sans craindre de piratage intensif.

- **Les articles de presse on-line**

Qu'en est-il de la mise en ligne d'articles de journalistes ? La cession sur support papier d'un article peut-elle entraîner implicitement sa numérisation sur le site du journal ?

L'ordonnance rendue par le Tribunal de Grande Instance de Strasbourg en faveur des journalistes des *Dernières nouvelles d'Alsace* en février 1998 apporte un premier élément de réponse. Le tribunal a jugé que le journaliste limite la cession de son droit d'auteur à une première publication et que la reproduction de l'œuvre d'un journaliste dans un autre périodique est soumise à autorisation.

Toutefois, il n'existe pas de règle générale et chaque journal invente sa propre solution. Ainsi, *Le Monde* a signé en 1996 avec les représentants syndicaux, un accord valable deux ans stipulant que l'autorisation accordée à l'éditeur pour la réutilisation d'articles sur Minitel, Internet ou CD-Roms se fait en contrepartie financière. Cependant, de nombreux éditeurs sont attachés à la notion de copyright à l'anglo-saxonne, qui leur permet de considérer le contenu d'un journal comme une oeuvre collective, susceptible d'être dupliquée sur différents supports. Le Syndicat de la presse parisienne et le Syndicat de la presse quotidienne régionale envisagent de demander au gouvernement de modifier la loi sur le droit d'auteur sans ce sens.

VII.6.b Les difficultés d'application de la législation liées à la dimension internationale d'Internet

Si les règles existent en matière de propriété intellectuelle, il est facile d'y contrevenir, notamment en raison de la mondialisation du réseau et de la coexistence de législations différentes, plus ou moins protectrices.

- **Les conflits de droits nationaux en matière de droit d'auteur**

En ce qui concerne les œuvres protégées, c'est l'auteur qui décide d'autoriser la mise à disposition du public de ses œuvres. Cependant, il faut admettre que dès qu'une œuvre d'auteur protégée est numérisée, et mise en circulation sur Internet, l'auteur perd tout contrôle sur sa diffusion. De nombreux groupes musicaux ont vu leurs morceaux les plus récents diffusés sans qu'ils en aient eu le contrôle.

L'auteur pourrait codifier son œuvre¹²⁶, pour permettre d'identifier son origine, son lieu de première publication, d'anticiper les utilisations possibles de son œuvre sur Internet afin de chiffrer la rémunération qui lui serait due. Mais comment circonscrire précisément le périmètre d'exploitation de l'œuvre sur Internet ? Il ne s'agit pas d'un acte simple. Comment appliquer les règles de droits d'auteur alors que le délit a lieu dans un pays qui a une législation différente ?

En effet, le cloisonnement des législations révèle d'importantes disparités. Certains Etats, en particulier les pays asiatiques, sont beaucoup plus laxistes en matière de poursuites des infractions. Un impératif est donc tout d'abord de définir la loi applicable, celle du pays émetteur ou celle du pays récepteur :

Dans le cadre d'un contrat, la loi applicable est celle que les parties ont adoptée, sous réserve des règles et conventions internationales.

En matière de responsabilité civile extracontractuelle, la loi applicable est celle du lieu où le fait dommageable s'est produit (lieu générateur du dommage ou lieu où il s'est produit).

La loi pénale applicable dépend des Etats. En France, selon le Code Pénal, « la loi pénale française est applicable aux infractions commises sur le territoire de la République. L'infraction est réputée commise sur le territoire de la République, dès lors qu'un de ses faits constitutifs a eu lieu sur ce territoire. ». Cependant, l'identification et l'incrimination d'un prévenu sont rendues difficiles en raison de l'utilisation de techniques d'anonymat. De plus, l'entraide répressive internationale est limitée par la règle de la double incrimination (qui existe en droit français), qui ne facilite la poursuite que pour des infractions répondant à une qualification pénale dans les deux Etats concernés. Dès lors, l'existence de paradis électroniques n'est pas à négliger.

¹²⁶ Des systèmes d'identificateurs uniques ont fait leur apparition, notamment le système de codecs DRM par Microsoft

La répression des fraudes est d'autant plus compliquée à mettre en œuvre qu'il est très difficile de localiser un serveur qui diffuse une œuvre sans autorisation. Tout le monde peut diffuser de chez lui, avec un ordinateur connecté à Internet et ensuite délocaliser son serveur. Ainsi, quand Pascal Barbraud s'est vu interdire la mise à disposition du livre du Docteur Gubler sur Internet, il a menacé de le mettre sur un serveur en République Tchèque.

Une solution est de réunir des pays autour d'une convention afin d'harmoniser certaines règles de droit. Ainsi une convention a eu lieu à Genève en Décembre 1996, pour s'interroger sur une éventuelle extension des droits d'auteur aux recopies successives d'une œuvre. Les opérateurs de réseau et de sites Internet désiraient que toute donnée publiée une première fois sur le Net devienne une donnée publique, compte tenu des recopies inévitables dues aux caractéristiques techniques d'Internet. A l'inverse, les producteurs étaient partisans d'une extension du droit de reproduction aux recopies même transitoires et même en l'absence de transfert chez l'utilisateur final.

- **Le droit des brevets : un niveau de protection hétérogène**

L'obtention d'une protection juridique sur un brevet est un leurre si, dans la pratique, il n'est pas possible de la faire respecter. Or beaucoup d'entreprises mettent en doute l'efficacité de la protection quand il s'agit de faire cesser et de réprimer la contrefaçon, en particulier dans certains pays d'Asie.

De ce point de vue, les règles de propriété industrielle en Europe sont jugées inadéquates, voire dissuasives, par de nombreuses entreprises. Leur évolution est nécessaire afin que des projets plus stratégiques puissent être soutenus. En effet, le brevet est par excellence l'outil de protection de l'innovation technologique, et sans doute celui que la France et ses entreprises utilisent le moins bien.

Afin de faciliter la diffusion et l'accès à l'information, un projet de réseau mondial d'information, baptisé WIPOnet, a été approuvé dernièrement par les Etats membres de l'OMPI. Le but de cette initiative est de créer des bases de données de qualité, pouvant être mises sur le réseau en toute sécurité par le biais de bibliothèques numériques permettant d'extraire des données relatives aux demandes internationales de brevets publiés. Il est ensuite envisagé de procéder au dépôt électronique des demandes internationales de brevet, permettant ainsi d'abaisser le coût du dépôt. A l'échelle européenne, l'Organisation européenne des brevets met à disposition près de 30 millions de documents de brevets publiés dans les 19 pays membres, afin de promouvoir l'innovation dans les entreprises européennes.

Au niveau national, une adaptation du système judiciaire paraît nécessaire. Il faut notamment examiner les possibilités de réévaluer les dommages et intérêts accordés aux victimes de la contrefaçon. En effet, en France, les indemnités sont fondées sur la réparation du préjudice subi par le breveté et non pas, comme aux Etats-Unis, sur la confiscation des bénéfices illicites réalisés par le contrefacteur. Il convient donc d'engager une réflexion sur ce thème, de façon à rendre les sanctions plus dissuasives. Aux Etats-Unis, la création en 1982 d'une cour d'appel unique au niveau fédéral pour les brevets et le triplement des sanctions financières infligées au contrefacteur «délibéré et conscient» ont beaucoup contribué à rendre confiance dans le brevet et à développer son utilisation.

- **Les conflits de droits nationaux et de régimes juridiques en matière de noms de domaines.**
 - un conflit de régimes juridiques

La coexistence des différentes procédures d'enregistrement de nom de domaine a été l'occasion de conflits importants entre entreprises, en raison de l'utilisation par des tiers de leur marque sans autorisation préalable. Ainsi, certaines entreprises se sont vues refuser l'emploi de leur propre marque pour leur site personnel. Elles ont été obligées de négocier d'importantes transactions financières afin de pouvoir employer leur marque, antérieurement enregistrée à l'Internic sous le suffixe « .com ». Ainsi, après le rachat de Digital et de son moteur de recherche appelé Altavista, Compaq a dû payer 3 millions de dollars à la société anglaise AltaVista Corporation. Cette possibilité de gains substantiels explique que des entreprises spécialisées mais aussi des particuliers, déposent des noms de domaine par pure spéculation. Ce piratage est rendu possible par le fossé existant entre le droit des marques et le droit des télécommunications. La règle dans le domaine des Télécommunications et appliquée par l'Internic, est celle du « premier arrivé, premier possesseur », alors que le droit des marques est fondé sur le principe de spécialité. Cela signifie qu'une marque est déposée et protégée pour désigner les produits ou services visés à son enregistrement. Le 12 mars 1998, le tribunal de Grande Instance de Paris a considéré, dans ce genre d'affaire, que reconnaître la thèse du « premier arrivé, premier servi » est incontestablement préjudiciable à la société la plus ancienne, qui se trouve ainsi dépossédée sur Internet de la dénomination sociale qui l'individualise dans l'ensemble de son existence et de ses activités, alors qu'elle jouit d'un véritable droit de propriété sur celui-ci.

- Un conflit de droits nationaux

La grande diversité des pratiques dans la gestion des différents noms de domaines (que ce soit au niveau de la nature de l'organisation responsable de la gestion, de la protection du droit des marques, des conditions d'enregistrement des noms de domaine) est à l'origine de conflits au niveau international.

Suite à l'augmentation de ce type de conflits, le NSI a élaboré une charte le 9 Septembre 1996, dont le but est de prévoir des modes de règlement des litiges concernant les noms de domaines. Elle prévoit la possibilité pour le titulaire d'une marque, de déposer une réclamation auprès du NSI, lorsqu'il aura été constaté l'enregistrement par un tiers d'un nom de domaine comparable à cette marque déposée. Le NSI adressera alors une mise en demeure au déposant du nom litigieux, le contraignant à fournir dans les trente jours un justificatif démontrant sa titularité sur la marque. Si l'organisme ou la société en question ne parvient pas à se justifier, le NSI lui demandera d'abandonner le nom de domaine. En cas de refus, le nom de domaine litigieux sera mis en attente et aucune des parties ne pourra l'utiliser jusqu'à ce qu'elles trouvent un terrain d'entente ou qu'un juge ait tranché l'affaire.

Une entreprise française pourrait bénéficier d'une telle procédure, dans l'éventualité où une firme étrangère enregistre un nom similaire à sa marque au sein de la zone « .com ». Cependant, le dispositif risque de s'alourdir, puisqu'il faudra soit diligenter une procédure dans l'Etat de l'entreprise contrevenante, soit obtenir la reconnaissance de la décision juridictionnelle française devant le tribunal étranger.

Au niveau de la zone «.fr», le contrôle effectué par le N.I.C-France permet d'éliminer les cas de fraude manifeste. L'enregistrement comme nom de domaine de la dénomination sociale d'une société concurrente sera impossible. Rappelons qu'en vertu de l'article L711-4 du Code de la propriété intellectuelle, il est interdit d'adopter un signe ou une marque, s'il existe un risque de confusion avec une dénomination sociale ou une enseigne déjà connue sur le territoire.

Cependant, des conflits peuvent survenir entre marques similaires désignant des produits différents, mais qui souhaitent utiliser le même nom de domaine sur le réseau. C'est pourquoi, une société désireuse d'enregistrer le nom de son site Web aura tout intérêt à déposer simultanément la marque correspondante. Il semble effectivement qu'en cas de litige, une marque déposée sera plus facilement prise en compte par un tribunal ou le N.S.I, plutôt qu'un simple nom commercial.

Quant aux titulaires de marques françaises, souhaitant une extension internationale, ils devront adresser une demande à l'I.N.P.I qui sera transmise à l'O.M.P.I. Il faut cependant rappeler, que les pays anglo-saxons et en particulier les Etats-Unis, n'adhèrent pas aux traités administrés par l'O.M.P.I.

Enfin, la Convention de l'Union de Paris du 20 mars 1883, où adhèrent cette fois les Etats-Unis, a mis en place un mécanisme de priorité, autorisant le déposant d'un enregistrement national, à procéder à des dépôts dans les autres pays membres, dans un délai de six mois.

On constate que le système actuel de gestion des noms de domaines sur Internet ne pourra satisfaire longtemps tous les titulaires de marques, ou les entités désireuses de protéger leur dénomination sociale. *« La création d'une instance réellement internationale devient une nécessité, On peut imaginer une structure dépendant de l'OMC afin de régler ce genre de problèmes »* estime Pierre Rebol, président de l'Electronic Business Group.

Les divers problèmes d'application qui viennent d'être évoqués peuvent néanmoins être résolus, tout au moins en partie. La plus grande des difficultés réside dans le caractère transfrontalier d'Internet face au morcellement des législations. Plusieurs solutions sont à envisager, qu'elles soient d'ordre techniques (cryptographie), juridiques, politiques (coopérations internationales) ou économiques (autorégulation).

- **La diversité des solutions**

Le nouveau mode de communication qu'Internet introduit à pour particularité essentielle un accès à l'information grandement facilité. Il s'affranchit des contraintes de distances et de temps, des frontières, et offre à tous un égal accès à l'information qui constitue sa richesse.

Le droit de la propriété intellectuelle, qui vise à réglementer la création et la transmission des œuvres de l'esprit doit donc se mettre en accord avec ces nouvelles facilités. Il appartient de même à la technique de faciliter l'application des règles déjà existantes, qui, on l'a vu, peuvent dans la plupart des cas s'appliquer. Cependant, le caractère transfrontalier du réseau implique une coopération internationale, non seulement des états, mais aussi des utilisateurs, qui doit déboucher sur une harmonisation des règles.

La conciliation entre la facilité de communication offerte par le réseau et les principes de la propriété intellectuelle devra donc se faire par le biais d'une coopération internationale.

- **La conciliation entre la facilité de communication et les principes de la propriété intellectuelle.**

Les nouvelles facilités de communication offertes par le réseau provoquent ou sont susceptibles de provoquer de nombreuses atteintes au droit de la propriété intellectuelle, mais, le réseau des réseaux offre aussi des opportunités culturelles très importantes. Il paraît donc souhaitable de concilier les principes de la liberté d'information et de la liberté d'expression, qui sont à la base du développement d'Internet, avec les principes du droit de la propriété intellectuelle, et notamment la protection des auteurs. Le réseau doit donc s'adapter au droit, et le droit doit s'adapter au réseau.

VII.6.c Le réseau doit s'adapter au droit

Comme nous l'avons déjà évoqué, la numérisation de l'information et la disparition du support matériel ont pour conséquence une nouvelle facilité de circulation, de modification, d'altération de l'information qui s'offre à tous. C'est ainsi que, d'une part, elle peut être facilement copiée ou modifiée, ce qui entraîne de nombreuses atteintes potentielles au droit d'auteur, concernant le respect de l'intégrité de l'œuvre, mais aussi le droit de reproduction. D'autre part, le développement du télétravail et des transmissions de données entre les différents sites d'une entreprises entraîne un accroissement considérable du risque de vol de données et donc d'atteintes à la propriété intellectuelle. Ainsi, de nombreuses entreprises utilisent des ordinateurs reliés au réseau dans le cadre de leurs programmes de Recherche & Développement ou de conception de nouveaux produits, ce qui peut conduire à des litiges concernant le droit de la propriété industrielle et plus particulièrement le droit des brevets.

Par exemple, la firme américaine Boeing, a conçu son dernier avion à partir de plusieurs bureaux d'études distants de plusieurs milliers de kilomètres et reliés entre eux par le réseau.

Pour encourager le développement des activités économiques, par la protection de la propriété intellectuelle, le réseau se doit donc d'offrir des solutions techniques de sécurité concernant la transmission et le stockage des informations, mais aussi l'authentification des documents. On peut penser, par exemple, à l'utilisation de la cryptographie RSA (Rivest Shamir Adleman) aussi nommée cryptographie asymétrique. Celle-ci fonctionne avec deux clés : une clé public et une clé privé. Les deux sont utilisées pour le cryptage, et la clé privée sert à décrypter. Ce système permet aussi bien de sécuriser les données (à l'instar des autres systèmes de cryptage), mais aussi d'authentifier les documents, éventuellement grâce au concours de serveurs de clés publiques, en créant une véritable signature électronique. Cependant, la cryptographie n'est pas encore totalement libéralisée. Il paraît aussi souhaitable de conserver une certaine liberté de communication. Le réseau, par l'avancée technologique qu'il apporte, doit donc aussi contraindre le droit à s'adapter.

VII.6.d Le droit doit s'adapter au réseau

Le droit de la propriété littéraire et artistique est assorti de tempéraments visant à promouvoir la liberté d'expression, liberté fondamentale, et à favoriser la circulation de l'information. C'est ainsi qu'en France, le Code de la Propriété Intellectuelle, dans l'art. L122-5, définit des tempéraments au droit d'auteur, tels que, la copie à usage privé. Il faudrait donc apporter de nouveaux tempéraments aux droits d'auteurs, plus adaptés à la nature du réseau. (Pour éviter des problèmes tels que l'assignation de propriétaires de pages Web personnelles qui reprennent des œuvres protégées, souvent sans savoir qu'ils sont dans l'illégalité). En effet le concept de copie à usage privé est peu adapté au monde digital car, d'une part, il permet le téléchargement et donc offre une grande facilité de duplication, et d'autre part, il introduit une difficulté dans la distinction entre privé et public (On peut penser au procédé des listes de diffusion).

Il paraît donc logique d'apporter des tempéraments au droit d'auteur qui seraient définis selon l'usage des œuvres. Ainsi, le droit anglo-saxon autorise la reproduction des œuvres protégées par le copyright si elle ne nuit pas à l'auteur. Cette exception d'usage loyal (ou fair use) implique la possibilité de faire des copies à usage privé pour le propriétaire de l'œuvre, mais aussi pour d'autres personnes, si cette copie est utilisée à des fins de recherche ou d'enseignement. Cette exception fait l'objet d'une jurisprudence au cas par cas, ce qui la rend difficile à systématiser. Néanmoins, le critère de l'usage non lucratif peut souvent induire son application. En outre, le développement du concept de shareware, souvent utilisé pour les logiciels, introduit une nouvelle exception. Dans ce cas, l'utilisateur peut se servir d'un programme pour l'essayer et ne doit rétribuer l'auteur que s'il s'en sert réellement. Il s'agit donc d'une nouvelle exception au droit d'auteur qui s'est développée « naturellement », et avec l'accord des auteurs.

On peut aussi regretter que la jurisprudence française dans le cadre de l'affaire Microfor / Le Monde qui autorisait les abstracts et résumés d'œuvres d'autrui n'ait pas été reprise par la directive européenne de 1996 sur les banques de données.

Cette question reste donc en suspens et nécessitera dans l'avenir une réponse adaptée au nouvel environnement, d'autant plus que les textes internationaux sont flous. En particulier, la convention de Berne, laisse la plupart du temps aux états membres, la faculté d'adopter des dispositions concernant les tempéraments à apporter aux droits d'auteur.

VII.6.e La nécessité d'une entente internationale

Le réseau Internet a souvent été considéré comme une zone de non droit parce que l'application de dispositions légales préexistantes se heurte à de multiples difficultés, notamment en raison de son caractère international et la multiplicité de ses acteurs. C'est ainsi qu'un autocontrôle semble préférable au système classique des réglementations législatives contraignantes. Celui-ci doit s'appuyer sur les règles de droit déjà en vigueur. Pour le mettre en place d'une manière efficace, une coopération internationale des états est indispensable, de manière à unifier les règles.

- **L'autorégulation**

L'autorégulation vise à obtenir de la communauté des internautes une application efficace des principes de la propriété intellectuelle. Elle s'adresse non seulement aux titulaires des droits, mais aussi aux utilisateurs et aux intermédiaires. Il faut donc mettre en place des institutions propres au réseau, qui ont pour but l'information et la responsabilisation des acteurs, ainsi que la prévention et le règlement des litiges.

- **Information et responsabilisation**

Tout d'abord, la création d'une charte de l'Internet, devant être respectée par tous les acteurs du réseau (producteurs de contenu, intermédiaires) semble souhaitable. Elle permettrait ainsi de clarifier les règles applicables et conduirait les sites hébergeant des pages Web à exiger de leurs clients le respect des principes de la propriété intellectuelle. Diverses propositions ont été déjà formulées, comme celle de la commission Beaussant qui a remis en mars 1997 au gouvernement une proposition de charte de l'Internet, celle de l'Association canadienne des Fournisseurs d'Internet (A.C.F.I) ou encore celle de l'association des providers britanniques (I.S.P.A.).

- **Prévention**

La création de dispositifs visant à prévenir les conflits par le biais d'une harmonisation des procédures est essentielle. Il faut donc évoquer les récentes initiatives que sont l'ICANN et le réseau mondial de l'OMPI. L'ICANN est une structure chargée d'administrer la politique de nommage des sites Internet¹²⁷. Il lui revient d'harmoniser les règles de nommage avec le droit international des marques de manière à éviter les pratiques de « cybersquatting ». Le réseau mondial d'information de l'OMPI doit, lui, relier les différents offices de propriété intellectuelle, et permettre à terme le dépôt électronique des demandes internationales de brevets. Ces deux structures devraient permettre de résoudre les problèmes touchant la propriété industrielle (droit des marques, droit des brevets). En outre, la création d'un observatoire international du réseau, à l'instar du rôle joué en France par l'union des fabricants (UNIFAB) peut permettre de déceler les atteintes à la propriété intellectuelle.

¹²⁷ Cf. "Les Enjeux liés à l'information".

- **Arbitrage**

Une instance d'arbitrage permet d'éviter les instances judiciaires. Cela permet un règlement non conflictuel des litiges mais surtout, évite les conflits entre les lois des différents pays, ce qui est particulièrement appréciable en matière de droit appliqué à Internet. C'est ainsi que dans le cadre de l'OMPI, un mécanisme de règlement accéléré des litiges a été institué auprès de la commission INTERDEPOSIT. Cette procédure, dite de « médiation et d'arbitrage en ligne », a vocation à régler le plus rapidement possible les différends survenus entre internautes en matière de droits de propriété intellectuelle et se déroule en une phase de médiation et une phase d'arbitrage. On peut aussi citer l'initiative du Centre de Recherche en Droit Privé de l'université de Montréal qui a développé un projet expérimental de résolution des litiges dans le cyberspace appelé cybertribunal, ainsi que le « virtual magistrate project », développé par L'A.A.A., une association d'arbitrage américaine.

Mais la mise en place et la pérennité de ces institutions ne peut se faire que par la voie d'une coopération au niveau internationale.

VII.7 La coopération internationale

La coopération doit être mondiale, mais il appartient avant tout aux pays européens de se mettre d'accord entre eux.

VII.7.a Au niveau européen

Dans le domaine de la propriété industrielle, le droit communautaire a d'ores et déjà institué des processus de dépôt de marques et de brevets à l'échelle de l'union européenne. Ainsi, l'Office des brevets de Munich délivre des brevets dont les effets s'étendent à tous les pays de l'Union Européenne, l'office d'harmonisation d'Alicante effectuant un travail similaire dans le domaine des marques. Il faut donc systématiquement encourager les dépôts de marques et brevets européens afin de remplacer à terme (au moins pour ce qui touche à Internet) les offices nationaux par les deux institutions précédemment citées. De même, il paraît souhaitable de développer les projets de dépôt en ligne tels que EASY et MIPEX. Par contre, dans le domaine de la propriété littéraire et artistique, la législation communautaire présente encore des lacunes. Néanmoins, la commission travaille à une harmonisation des législations des états membres et a publié le livre vert sur le droit d'auteur et les droits voisins dans la société de l'information en 1996, ainsi qu'un projet de directive en 1997. De plus la directive de 1996 sur les banques de données aborde certains aspects du droit d'auteur.

L'Union Européenne semble avoir réellement pris conscience de l'importance d'une législation communautaire sur la propriété intellectuelle, ce qui peut et doit lui donner une bonne position dans les négociations qui auront probablement lieu au plan international.

VII.7.b Au niveau mondial

La protection de la propriété intellectuelle n'a cessé de s'affirmer depuis la convention de Berne dans les pays industrialisés, et depuis les accords ADPIC conclus dans le cadre du GATT en 1993, dans les pays en voie de développement. Mais les problèmes soulevés par l'avènement d'Internet et de la société de l'information vont probablement conduire à de nouvelles négociations internationales sous l'impulsion de l'OMPI. C'est ainsi que se met en place l'ICANN, solution aux problèmes de nommage. Mais on a vu que de nombreux problèmes ne trouvaient de solution qu'au niveau international. De nouvelles négociations auront lieu, donnant probablement naissance à de nouvelles autorités de régulation

Les autres sources d'information sur le Web

Parce qu'il n'y a pas que les moteurs de recherche dans la vie et que les substituts en matière de recherche d'information ne manquent pas, nous dresserons ici un panorama des sources d'information accessibles sur le web. Nous présenterons les blogs, les banques de données payantes, les forums, les annuaires spécialisés, les services de pîges, la presse en ligne, la tendance du podcasting, etc. nous reviendrons également sur le projet de librairie en libre accès de Google : Le projet Google Print.

VIII. Les autres sources d'information sur le Web

Outre les moteurs de recherche, de nombreuses sources d'information sont à la disposition des internautes en fonction de leur bourse et de leur sensibilité aux nouvelles tendances.

VIII.1 Blogs (*joueb, Weblogs, Webillards*)

Les blogs sont à la mode. Tout le monde en parle, tout le monde affirme en posséder un, et ceux qui n'en ont pas, n'ont apparemment rien compris.

Les blogs sont apparus en 2003 et ils se sont multipliés et déclinés dans tous les secteurs : affaire, culture, politique. En 2005, la société d'études Perseus en dénombrait 30 millions à l'échelle mondiale, dont environ 3,5 millions en France. Le nombre de blogs au niveau mondial devrait atteindre selon les estimations, 50 millions d'unités d'ici à janvier 2006.

L'attrait du blog réside dans le fait qu'il revient à consulter le journal intime de son auteur. En effet un blog permet à chacun de devenir éditorialiste en freelance. Le blog offre à chacun la possibilité de partager ses expériences, ses connaissances, ses interrogations, avec le reste de la communauté. On trouve des blogs sur à peu près tous les sujets.

Les auteurs sont généralement des experts (mais pas uniquement) qui partagent toutes les connaissances qu'ils ont accumulés sur un sujet. L'autre avantage des blogs est leur maintenance relativement aisée. En effet ils sont mis à jour de façon continue (beaucoup plus rapidement que sur un site standard) et les réactions des visiteurs y sont consultables.

Bien entendu, étant donné la croissance rapide du nombre de blogs, le problème se pose encore une fois de disposer d'outils permettant de s'y retrouver, et c'est pourquoi se sont développés des moteurs de recherches spécialisés dans les journaux en lignes.¹²⁸

VIII.2 Les banques de données

Avec la culture du gratuit suggérée par le Web, le réflexe est vite conditionné de se jeter sur un moteur de recherche pour satisfaire le moindre besoin d'information. Attention à ne pas céder à la paresse : la recherche d'information doit être s'inscrire dans une démarche réfléchie.

En effet, une recherche concernant un secteur donné, une entreprise, une recherche sur les moteurs renverra des résultats trop nombreux. En revanche une base de données payante donnera le résultat en quelques clics.

Le contenu des bases de données est filtré, validé, structuré. Pour trouver une base de données, paradoxalement, il faut consulter des répertoires spécialisés qui trient les bases de données professionnelles selon des critères qualitatifs. Attention ces sites ont deux défauts : la plupart sont en anglais (d'où l'importance de recouper les

¹²⁸ www.technorati.com, www.blogpulse.com, www.blogsearch.google.com

informations obtenues dans différentes langues) et sélectionnent en priorité des documents anglo-saxons. D'autres sites comme Dialog.com ou Lexisnexis.fr regroupent plusieurs bases que l'on peut acheter directement ce qui n'est pas le cas avec les répertoires. Ces bases sont classées avec les noms des éditeurs qui les commercialisent.

L'inconvénient des bases de données est le coût.

D'autre part, il faut maîtriser leur interface (généralement elles utilisent un système de recherche booléen) mais on peut demander à l'éditeur de la base une formation aux outils.

Exemples de bases : delphes.com, factiva.com, xerfi.com, kompass.fr

VIII.3 Forums (agora, ou newsgroups)

Les forums sont des lieux de discussion, où les internautes s'expriment en toute liberté, à chaud, sans se censurer, et c'est là leur principal intérêt.

Sur cette source d'information, on peut récolter des informations pointues, introuvables sur des pages Web classiques. Et ce d'autant que ces newsgroups sont ultraspécialisés. De manière générale, mieux vaut opter pour un forum régulé par un modérateur : il veille à la bonne tenue des échanges et débats et filtre les messages les plus fantaisistes. Consulter les annuaires et moteurs de recherche de forums comme www.rezoWeb.com, www.yahoo.groups.com, www.groups.google.com.

Le fait d'intervenir sur ces forums permet d'obtenir des réponses précises, introuvables autrement, c'est l'intervention de la communauté qui amène petit à petit la construction d'une réponse valable.

L'inconvénient de ces newsgroups est qu'il est possible, dès lors que l'on y est intervenu de renseigner les bases de données des moteurs de recherche de newsgroups.

Car certains serveurs ont constitué des bases de données de ces newsgroups et permettent de rechercher dans le texte intégral de toutes les interventions courantes et passées. Le problème est que certains de ces serveurs utilisent également comme entrée de leur index l'adresse e-mail des personnes qui sont intervenues sur les newsgroups. Si par exemple, vous utilisez beaucoup Internet et que vous êtes très ouvert sur le monde, vous posez sur les newsgroups des questions sur des sujets divers, vous participez aux débats, en politique, religion, musique. Ici ou là, vous êtes intervenu dans quelques polémiques sur des sujets qui à l'époque de l'intervention étaient d'actualité.

Grâce à ces moteurs de recherche de newsgroups, il est possible à partir de l'un des messages que vous avez envoyés, de récupérer toutes les autres interventions que vous avez faites sur tous les autres newsgroups et ainsi d'obtenir un profil assez net de vos centres d'intérêt. Il suffira après de lire chacune de vos interventions pour obtenir votre opinion détaillée sur l'ensemble de ces sujets.

Certes, si vous êtes intervenus, c'était pour être lu. Mais pas nécessairement pour que l'on puisse recouper vos différents messages. Ainsi, un employeur, peut par exemple, aisément obtenir des informations que vous ne lui auriez pas forcément donné sur sa demande et qu'il n'aurait pas été forcément légal de sa part de recueillir.

Intervenir sur un newsgroup est un acte réfléchi qui se situe dans un contexte : les messages s'effacent relativement vite des newsgroups, la quantité de newsgroups (plus de 15.000) et d'interventions pourrait vous donner à penser que les vôtres sont assez

anodines. Il est important de réaliser que ce n'est pas le cas et que tout comme le serveur de newsgroups peut tracer vos lectures de ces messages, des moteurs de bases de données accessibles à tous peuvent tracer vos interventions, les conserver aussi longtemps qu'ils le souhaitent et vous profiler à volonté. Ce qui représente aussi un trésor pour les services d'intelligence marketing.

VIII.4 Annuaire spécialisés

« Pour obtenir rapidement les adresses de référence sur un domaine d'activité, une profession ou un marché, rendez vous sur ces sites élaborés par des spécialistes.

Les annuaires sont des outils très complémentaires des moteurs de recherche. C'est particulièrement vrai lorsque vous explorez toutes les facettes d'un sujet sans savoir quels mots-clés indiquer. Ils permettent d'éviter de se retrouver avec deux millions de réponses en pareille circonstance si on utilisait un moteur de recherche.

Pas évident alors de repérer les sites de référence. C'est précisément la fonction d'un annuaire. Ils permettent d'accéder à des sites répartis en plusieurs catégories. Tous ces sites abordent de façon complète les questions qui vous intéressent. Logique car si les moteurs de recherche se contentent d'indexer automatiquement des mots-clés dans des pages Web (ce sont des robots), les annuaires sont administrés par des hommes qui recensent et reclassent manuellement les meilleurs sites. Mieux, les adresses sont généralement accompagnées d'un commentaire précisant leur intérêt.

Il faut préférer les annuaires spécialisés, à la différence des annuaires généralistes comme Yahoo ! Seuls les annuaires spécialisés se révèlent efficaces.

Ils présentent l'avantage de recenser des sites intéressants mais peu visités donc peu populaires, ignorés par les moteurs de recherche.

Il existe des annuaires d'annuaires comme www.enfin.com, www.lesannuaires.com, ou encore www.indicateur.com. »

VIII.5 Presse en ligne

Au début des années 1990, l'essor des NTIC a généré les premières divisions Internet des entreprises de presse. Dès 1994, une trentaine de titres était en ligne dans le monde. En France, le nouveau mode d'expansion a été occulté par les difficultés financières des sociétés. Rapports et avis se multipliaient (notamment rapport Detailleur, "L'évolution de la presse écrite dans la perspective des nouvelles technologies multimédias" en 1994; rapport Bourdier "La presse et le multimédia" en 1997), concluant unanimement que "l'avenir de la presse passe par ses développements sur le net et sur la télévision". Ce contexte a semblé-t-il précipité l'intervention des pouvoirs publics, une labellisation des projets a été mise en place à partir d'octobre 1995. Mais cette opération n'a offert qu'une place réduite aux professionnels. Etudes d'impact, investissements techniques, recrutement d'équipes ad hoc de journalistes et de juristes, hébergement et développement des sites... la mise en

page sur le réseau est toujours un challenge capitalistique majeur pour une entreprise de presse¹²⁹.

A la source de l'information, les agences de presse commercialisent un service d'articles de presse en continu. Mais la reprise mécanique de ce dispositif ne saurait constituer la vitrine Web d'un entrepreneur à la recherche d'une notoriété accrue. La concurrence débute en fait au niveau de l'agence et de l'architecture des sites.

Au départ simple reproduction sur le Web de l'édition papier, actualisée et en libre accès, l'édition électronique s'efforce désormais de définir une logique d'entreprise. La démarche exige notamment actualisation permanente et personnalisation du contenu (comme pour les blogs).

Sans services payants, bon nombre de sites seraient voués à disparaître. Pour les éditeurs de presse écrite, la première légitimation de cette formule est l'accès aux archives. Cette prestation est proposée depuis longtemps par le site "Lesechos.fr".

Pour autant, la multiplication des services en ligne ne semble pas priver la presse traditionnelle de son public potentiel. La pratique du Web aurait même un effet favorable à la lecture des magazines. Une récente étude de Nielsen/NetRatings montre qu'un utilisateur d'Internet sur cinq qui lit des journaux et magazines préfère s'orienter vers la version en ligne de cette presse papier. 21% exactement des habitués du Web fonctionnent donc ainsi, tandis que la majorité (72%) opte encore pour la version papier et que 7% alternent entre les deux versions. Il faut toutefois reconnaître qu'il est plus confortable pour l'oeil de lire un article sur le papier que sur un écran.

Gerry Davidson, analyste pour Nielsen/NetRatings, estime qu'il s'agit là d'une très bonne nouvelle pour la presse en ligne.

L'étude a été réalisée sur 9000 personnes de janvier à mars et montre que les hommes sont plus enclins à se tourner vers Internet comme source d'informations.

Observons aussi la vertu du "copyright" anglo-saxon qui permet une plus grande liberté d'action : le contenu d'un journal est une oeuvre collective susceptible d'être dupliquée sur différents supports. Le droit français a connu pour sa part une évolution différente. Le projet "Figaro.fr" avait été par exemple suspendu jusqu'à la signature d'un accord sur les droits d'auteur, négocié par les syndicats de journalistes à l'issue d'une décision de justice. Celle-ci estima, le 20 mai 2000, qu'une conciliation des droits sur chacune des contributions et sur l'ensemble était tout à fait réalisable. La mise en oeuvre d'une charte d'édition électronique a concrétisé cette évolution.

La presse en ligne est encore un média en devenir¹³⁰. Aussi un observateur averti pourra-t-il déplorer, de la part des éditeurs et des journalistes, l'exploitation insuffisante de l'interactivité. Ce dernier semblerait in fine avoir vocation à réaliser une communauté d'intérêts autour des journalistes (a debating society), à explorer de nouveaux modes de communication avec les lecteurs (new digital story-making) et à

¹²⁹ Source : <http://www.minefi.gouv.fr/Webmen/revuedeWeb/presseligne.html>

¹³⁰ L'offre de presse disponible en ligne est très large, notons que des annuaires de journaux et magazines de la presse francophone comme <http://www.giga-presse.com>, <http://www.journaux.fr>

donner la parole au public par la multiplication des blogs¹³¹, mais peut-on encore parler de journalisme ?

VIII.6 Services de piges (agents¹³², RSS¹³³)

Avec un logiciel utilisant la technologie RSS, vous pouvez demander à votre ordinateur d'aller chercher tout seul vos infos. La technologie du fil RSS évite à l'utilisateur d'avoir à consulter plusieurs fois par jour ses sites ou blogs préférés. Elle permet d'importer automatiquement sur votre ordinateur les informations qui vous intéressent. Le RSS comme le HTML est un langage informatique utilisé par de nombreux sites Web. Avantage : certains programmes spécialisés sont capables d'aller chercher des articles au format RSS et d'en afficher le contenu sur votre ordinateur. Ces lecteurs RSS comme Alertinfo ou ActiveRefresh sont téléchargeables gratuitement sur le Web. La syndication RSS est matérialisée sur les sites par des étiquettes oranges ou bleues portant la mention RSS ou XML. Copiez ce lien à l'aide de votre souris, puis collez le dans la rubrique « add channel » de votre lecteur RSS.

Les informations sont présentées de façon concises (un titre, les premières lignes, une couleur si le texte est récent) il suffit de cliquer pour en savoir plus.

VIII.7 La baladodiffusion ou « podcasting »

- **La nouvelle forme de Citizen band (« cibie »)**

Le vocabulaire du Web s'enrichit tous les jours avec les nouveaux moyens qui sont mis de l'avant pour séduire les publics cibles de l'Internet. Il y a eu l'ajout récent du mot blogue au dictionnaire Le p'tit Larousse et il faudra sérieusement considérer la baladodiffusion d'ici peu avec la popularité croissante de cette forme de diffusion. Chez les anglophones, on parle de la baladodiffusion sous le nom de podcasting.

La baladodiffusion permet aux utilisateurs de s'inscrire à un flux (feed en anglais) et ainsi de récupérer de nouveaux fichiers audio automatiquement. Le podcasting se distingue des autres moyens de diffusion d'informations par son utilisation du format RSS.

D'après les explications fournies sur le site Wikipédia, le podcasting est différent du broadcasting et du Webcasting en ce qu'il diffuse du son, non pas par un mécanisme

¹³¹ News Sites Repeat Mistakes Of the Past, Editor and Publisher, 10 avril 2002

¹³² Exemple de Google.news, service de pige sur Internet, et Google.Alerts, agent automatique qui envoie les actualités concernant un sujet donné dans votre boîte mail. Les alertes existent aussi sur certains sites dédiés à l'emploi comme Monster.fr.

¹³³ RSS: « Rich Site Summary » ou « Really Simple Syndication ». Le RSS est un format informatique permettant la « syndication », ou l'agrégation, de contenu », c'est-à-dire la possibilité de publier automatiquement sur un site Web des informations issues d'un autre site : dernières nouvelles, articles, nouveautés...

centralisé qui enverrait un flux vers ses auditeurs, mais par l'action des auditeurs qui vont aller chercher eux-mêmes les fichiers audio (pull au lieu de push).

Les francisations baladodiffusion et baladodiffusion ont été proposées fin octobre 2004 par l'Office québécois de la langue française.

La prolifération des lecteurs MP3 permet donc à un lot considérable de personnes de se procurer des émissions en podcasting.

En avril dernier, la Pew Internet and American Life Project a estimé à 22 millions le nombre d'Américains possédant un lecteur MP3. 29 % d'entre eux auraient téléchargé des baladodiffusions. Le téléchargement de spectacles, de chroniques de vin et d'émission de radio donne à l'auditeur la liberté d'écouter n'importe où, n'importe quand, une émission qu'ils auront téléchargée sur le Web.

Si la tendance se poursuit en ce sens, plusieurs corporations pourraient être tentées de séduire leur clientèle par le biais de ce genre de diffusion. La baladodiffusion est donc en voie de se tailler une place de choix dans le cœur des internautes¹³⁴.

VIII.8 Les encyclopédies collaboratives en ligne

Elles ont préexisté aux blogs, et tout comme eux, elles sont enrichies bénévolement par des milliers de contributeurs. Le site de référence est Wikipédia, avec 1,3 million d'articles écrits par 36 000 passionnés dans 212 langues (des plus courantes aux plus rares, telles le kurde ou...l'espéranto).

Créé en janvier 2001 par l'Américain Jimmy Wales, un éditeur de sites érotiques fasciné par les logiciels libres, comme « Linux » (conçu par des programmeurs bénévoles), le site comptait 20 000 articles au bout d'un an, et près de 100 000 l'année suivante. Parallèlement aux encyclopédies généralistes comme Wikipédia¹³⁵, des dizaines de sites spécialisés dans le cinéma, les animaux ou le sport fonctionnent selon le même principe¹³⁶.

L'intérêt de ces encyclopédies est la possibilité d'obtenir des informations sur un sujet général et/ou spécifique (comme pour lorsqu'il est préférable de choisir de consulter un annuaire Web).

Wikipédia contient 85 000 articles francophones (contre 30 000 pour l'Encyclopaedia Universalis version papier) et 500 000 en anglais¹³⁷.

Sur Wikipédia, n'importe quel visiteur peut modifier les définitions, et les contenus sont actualisés en temps réel. L'inconvénient est bien évidemment, le fait que les articles ne sont pas toujours rédigés par des experts ou des spécialistes mais ils sont relus dans la mesure du possible par des modérateurs qui attribuent des labels de confiance aux articles. D'autre part, les lecteurs qui ne sont pas d'accord avec les définitions proposées, ont la possibilité de discuter leur point de vue avec l'auteur sur le forum de Wikipédia.

¹³⁴ Source : <http://www.acadie.net/chronique/contenu.cfm?identification=6977>

¹³⁵ Ou encore www.Webencyclo.com, et <http://fr.wictionary.org>

¹³⁶ Comme www.imdb.com, www.discogs.com, www.gracenote.com/music, www.aquabase.org, www.wikif1.org, www.wikitavel.org

¹³⁷ Source : Management, Avril 2005.

Les chercheurs eux-mêmes sont partagés lorsqu'ils jugent de la qualité de Wikipédia, mais les inconditionnels de l'encyclopédie collaborative répondent aux détracteurs de Wikipédia que les erreurs sont rares, grâce, justement, au contrôle exercé par la communauté de contributeurs. Encore une fois le bon sens et la responsabilité de chacun face à l'information reçue sont de rigueur.

VIII.9 Le projet Google Print

Le 14 décembre 2004, Google annonçait au monde entier sa ferme intention de créer le premier service numérique accessible à tous, permettant de consulter les livres libres de droit sur Internet, soit 15 millions d'ouvrages.

Le projet Google Print consiste à scanner les pages de ces 15 millions d'ouvrages, et de les mettre en ligne, le tout, en a peu près 6 ans ! Le coût estimé de cette opération serait entre 150 et 200 millions de dollars.

Au-delà de la démarche culturelle, il est à noter que ce nouveau service fait partie d'une stratégie de globalisation des compétences de Google, visant à rendre le prestataire incontournable. D'autre part, il s'agit de mettre en ligne le contenu intégral des ouvrages libres de droits, ceux étant protégés par des copyrights seront présentés dans un contexte proche d'un esprit « publi-rédactionnel ». Ces derniers ouvrages seraient donc mis en avant sur Google Print de manière à ce que l'utilisateur puisse se rendre sur les sites des éditeurs de ces ouvrages.

Dans le cadre de ce projet, Google collabore avec les bibliothèques des universités de Stanford, Michigan, Harvard, ainsi qu'avec la New York Public Library. Cependant, cette collaboration ne s'étend pas que sur le seul territoire Nord américain, puisque la Bodleian Library, rattachée à la prestigieuse université d'Oxford en Angleterre, serait également associée.

Aussi, à la lecture simple du communiqué officiel de Google, l'utilisateur moyen du moteur de recherche pourrait se contenter de penser que cela ne concerne que des ouvrages écrits en langue anglaise.

Le président de la Bibliothèque Nationale de France (BNF), Jean-Noël Jeanneney, a déclaré qu'il n'était pas hostile à l'entreprise de Google, mais qu'il était troublé de l'orientation de la sélection des œuvres scannées par la société, dans la mesure où un grand nombre « de titres européens » figuraient dans la liste des ouvrages mis en ligne par Google Print¹³⁸.

Le directeur de la BNF s'est donc basé sur un postulat : « ...Leur sélection, leur hiérarchisation dans les listes vont être définies forcément à partir d'un regard singulier : celui de l'Amérique [...] La production anglo-saxonne sera inévitablement surévaluée. Le miroir américain sera le prisme unique. » .

Offrir un libre accès à tous à ces lectures est un rêve qui permettrait de réduire les inégalités culturelles, mais encore une fois les questions commerciales viennent biaiser la démarche et ce sont bien la situation prévisionnelle de monopole de la société souhaitant développer cette offre et leur vision « étasunienne » de l'information qui dérangent le reste du monde et en particulier les nations qui rayonnent à l'échelle du monde par leur patrimoine culturel.

Ce que souhaitent ces nations, c'est que l'information du monde demeure en flux visible sous différents angles. Une vision multipolaire, versus une vision unique, suggérée.

¹³⁸ Qui était disponible sur Internet dans sa version bêta en mai 2004.

Jean-Noël Jeanneney s'interroge également sur le financement du projet. Il indique que Google reste "flou" à ce propos. Or la crainte du président de la BNF est de voir une pression de la publicité et des firmes multinationales "le page Rank". Il soulève également le problème des droits d'auteurs et d'éditeurs qui complique bien la vie de Google.

Reste que pour mener à bien le projet de bibliothèque virtuelle européenne il y a encore quelques murs à faire tomber et quelques dangers à contourner. Par exemple, les difficultés à aboutir de l'union du fait de son enlèvement administratif. Mais aussi, ce qui pourrait être perçu au sein de l'union et dans le monde comme "l'arrogance française". Enfin, le dernier défi est principalement celui de la technologie à mettre en place notamment en ce qui concerne l'organisation des métadonnées.

Les Enjeux liés à l'information

Pourquoi l'information est elle un sujet sensible ?

En allant voir chez nos voisins moins libéraux, quelles différences faut-il noter dans la liberté individuelle face à un espace censé en être le symbole de la libre-expression ?.

Qui contrôle aujourd'hui l'information se trouvant sur le web et ce contrôle est-il légitime ?

Pourquoi cette question du contrôle des contenus Web et donc de l'information est-elle créatrice d'incidents diplomatiques ?

Identifions les enjeux liés à ce contrôle de l'information et envisageons des solutions pour l'avenir...

IX. Les Enjeux liés à l'information

IX.1 Tous égaux devant l'information ?

Cette surabondance d'information dont nous parlons depuis le début de ce mémoire permet-elle à tous d'être informés ?

Répondre par l'affirmative à cette question permettrait de conclure que nous sommes tous égaux devant l'information...malheureusement nous ne le sommes pas.

Nous n'avons pas tous la même facilité d'accès à l'information.

Reporters sans frontières, une organisation pour la défense de la liberté de presse, estime dans son rapport intitulé « Les ennemis d'Internet, les entraves à la circulation de l'information sur Internet¹³⁹ », que 59 pays font obstruction à la circulation de l'information sur Internet. Si le Canada ne fait pas partie de cette liste, la Corée du Nord, l'Afghanistan, la Chine, la France et les États-Unis y figurent, ces pays souhaitant restreindre, voire contrôler à différents degrés l'accès et la circulation de l'information jugée contraire à leurs intérêts ou à leurs principes.

IX.1.a Corée du Nord et Afghanistan

Ce rapport nous apprend qu'en Corée du Nord, Internet n'est disponible que pour le gouvernement, qui diffuse sa propagande à partir de sites officiels hébergés à l'étranger. Certaines organisations pour la protection des droits de l'homme utilisent aussi cette technique pour diffuser de la même façon une information tout à fait différente de celle du gouvernement.

La même situation prévaut en Afghanistan où, bien qu'il n'y ait aucun fournisseur d'accès à Internet, il a déjà été possible d'accéder au réseau des réseaux par le Pakistan ou par des téléphones satellites. Toutefois, en août dernier, on apprenait que le mollah Mohammed Omar interdisait tout accès à Internet... à une exception près: son bureau, qui ne peut être utilisé que par lui et son gouvernement. Le ministère « pour la promotion de la vertu et la prévention du vice » serait chargé de l'application de ce décret.

IX.1.b Chine

Des spécialistes de l'organisation Cyber-rights and Cyber-liberties disent de la réglementation chinoise qu'elle met Internet en « garde ouverte ». Ils soutiennent que le gouvernement souhaite tirer profit des avantages économiques d'Internet tout en évitant les conséquences que pourrait avoir cet outil de communication sur la stabilité politique.

En 1997, plusieurs instruments législatifs ont été approuvés par le Conseil d'État afin d'exiger des fournisseurs de services qu'ils établissent mensuellement des rapports sur le profil des utilisateurs d'Internet et qu'ils collaborent avec le « bureau de la sécurité publique » lors d'investigations liées à Internet. Mais surtout, ces instruments tiennent

¹³⁹ publié le 28 février 2001, à Paris,

<http://www.barreau.qc.ca/journal/frameset.asp?article=/journal/vol33/no20/surlenet.html>

http://www.rsf.org/rubrique.php3?id_rubrique=432 (Internet sous surveillance 2004)

les fournisseurs de services responsables des violations qui pourraient y être faites. C'est pourquoi ces derniers surveillent tout ce qui s'y passe, allant, dit-on, jusqu'à effacer les messages susceptibles de contrevenir à la liste des usages acceptables sur Internet établie par le gouvernement: il est notamment interdit d'inciter à la division du pays et de tenir des propos contre les organes de l'État ou du système socialiste.

De plus, en l'an 2000, trois lois ont été adoptées sur les bases des instruments législatifs de 1997. D'abord, les entreprises étrangères désirant investir dans le secteur d'Internet doivent obtenir l'autorisation du ministère de « l'industrie de l'information »; ensuite, les sites de nouvelles ne peuvent que publier de l'information préalablement censurée par le gouvernement; et finalement, la troisième loi criminalise les actes reliés à Internet. C'est ainsi que le « cyber-crime » et la « cyber-dissidence » ont été instaurés et consistent à diffuser des rumeurs, à diffamer ou à publier de l'information nuisible au gouvernement chinois, au système socialiste ou incitant à la division du pays.

IX.1.c France et États-Unis

Sans être aussi encadrant, la France a senti le besoin de légiférer pour mieux contrôler l'information sur Internet. Par un amendement à la loi française sur la liberté de la communication, les hébergeurs de sites Web ont depuis juin 2000 l'obligation de détenir et de conserver les renseignements permettant l'identification de toute personne hébergeant un site chez eux et de les communiquer au juge qui leur en ferait la demande. Les créateurs de sites doivent donc décliner leur identité et leurs coordonnées à l'hébergeur, ce qui est vivement dénoncé par les défenseurs de la vie privée.

Les États-Unis sont également actifs avec, ponctuellement, de nouveaux projets de loi. Les associations américaines pour la défense des libertés civiles soutiennent toutefois que le Premier amendement de la Constitution américaine garantit la liberté d'expression et elles exercent des pressions pour empêcher qu'Internet ne soit étouffé par la législation. Malgré tout, en décembre 2000, la puritaine société américaine se dotait d'une loi rendant obligatoire l'utilisation de logiciels de filtrage dans les écoles, les bibliothèques ainsi que dans les autres lieux publics ayant l'accès à Internet afin de bloquer l'information contenant notamment le mot « sexe ».

Et c'est sans compter qu'un système de surveillance du réseau à la solde du FBI, Carnivore, peut être utilisé sans mandat. À l'aide d'une boîte noire placée chez les fournisseurs d'accès à Internet, ce système intercepte les communications électroniques des personnes soupçonnées, à tort ou à raison, de compromettre la sécurité ou les intérêts du pays. Si ce système ne contrôle pas directement l'information qui transite sur Internet, il peut dissuader les citoyens de l'utiliser.

Bien que les entraves à la circulation de l'information sur Internet, perpétrées dans les 59 pays visés par le rapport de Reporters sans frontières, ne soient pas toutes du même ordre, le phénomène mérite que l'on s'en inquiète.

IX.2 A qui revient la responsabilité de contrôler l'information Web ?

Les Etats-Unis sont le berceau historique de la toile.

Créé en 1998 au terme de longues négociations menées par le vice-président américain Al Gore avec toutes les parties prenantes: chercheurs, industrie des télécoms, fabricants d'équipements, fournisseurs de contenus, administrations diverses, et le fameux professeur Jon Postel, l'ICANN est une organisation internationale sans but lucratif dont le rôle premier est d'allouer l'espace des adresses de protocole Internet (IP), d'attribuer les identificateurs de protocole, de gérer le système de nom de domaine de premier niveau pour les codes génériques (gTLD) et les codes nationaux (ccTLD), et d'assurer les fonctions de gestion du système de serveurs racines¹⁴⁰. Par le contrôle qu'elle exerce sur l'affectation des noms de domaines de premier niveau, l'ICANN dérive en pratique un droit de délégation sur la vente des noms de domaines à différentes organisations, comme VeriSign pour les domaines .com et .net ou l'AFNIC pour le domaine .fr.

Sa compétence est mondiale et ses décisions s'imposent de fait aux Etats, alors même qu'elle est de droit californien, se trouve soumise de ce fait à l'Attorney général (ministre de la Justice) de cet Etat et que le département du Commerce américain exerce un droit de veto sur ses décisions.

L'ICANN est dirigé par un directoire composé actuellement de 18 directeurs et elle supervise le fonctionnement des treize «serveurs racines» (root servers) sur lesquels repose l'architecture de l'Internet à l'échelle mondiale. Pour des raisons historiques, dix de ces machines se trouvent aux Etats-Unis, une autre est installée à Tokyo et les deux dernières en Europe, à Amsterdam et Stockholm.

Le rôle de l'ICANN est régulièrement remis en question, notamment à cause de ses liens avec l'administration américaine. L'ICANN a en effet été fondée suite à une directive du Département du Commerce américain et l'ICANN fonctionne toujours sur la base d'un mémorandum avec ce ministère.

De nombreux pays aimeraient que la fonction dévolue actuellement à l'ICANN soit prise en charge par un organisme dirigé par l'ONU. L'ONU elle-même a indiqué avoir des projets dans ce domaine, mais l'administration Bush s'y oppose fortement.

D'autres personnes préféreraient une autre organisation sous forme coopérative qui ne soit, ni l'ONU (qui est intergouvernementale et où la société civile n'est pas représentée), ni l'ICANN, qui est purement américaine.

Lors du dernier Sommet Mondial sur la Société de l'Information (SMSI) qui s'est tenu à Tunis le 15 novembre 2005, et dont l'objectif était de résorber la "fracture numérique" entre pays du Nord et du Sud, un accord sur une évolution de la gestion de l'Internet est intervenu¹⁴¹. Cet accord a permis d'éviter une rupture entre les Etats-Unis – hostiles à tout contrôle international – et le reste du monde, ont annoncé des

¹⁴⁰ Source : <http://www.icann.org/tr/french.html>

¹⁴¹ Source : Contrôle d'Internet: accord à Tunis entre les Etats-Unis et les autres pays, AFP, 15/11/2005

négociateurs SMSI de Tunis. L'accord est intervenu tard, quelques heures avant l'ouverture officielle du sommet, au terme de trois journées d'intenses discussions à Tunis. Le principal négociateur américain, David Gross, s'est félicité du compromis, qui prévoit la création d'un "forum" international destiné à discuter des questions relatives à l'Internet, sans toucher au statut actuel de l'Internet Corporation for Assigned Names and Numbers (Icann), la société californienne chargée d'attribuer les "noms de domaine" tels que .com, .org, .fr ou .ch. Une grande majorité de pays participants auraient souhaité imposer une structure internationale au-dessus de l'Icann, qui est actuellement chapeauté par l'administration américaine. Le compromis apparaît proche d'une proposition de l'Union européenne, dont les positions étaient défendues par la présidence britannique.

La première réunion devrait avoir lieu au premier semestre 2006 en Grèce, Athènes ayant offert d'accueillir les travaux, et après convocation du secrétaire général de l'Onu, Kofi Annan. Le forum discutera de sujets aussi variés que le courriel indésirable (spam), les virus informatiques ou la cybercriminalité.

La conclusion est toujours la même : Washington s'oppose à toute volonté d'instaurer un contrôle international de la toile.

Washington s'opposait fermement à une mise sous tutelle de l'Icann par un organe de l'Onu, estimant que cela donnerait à des pays qui réprime la "cyberdissidence" un pouvoir indu sur le net. Cette crainte était partagée par l'organisation de défense de la liberté de la presse Reporters sans frontières (RSF).

En admettant que l'on aboutisse un jour à la création de cette instance de contrôle internationale, supra-américaine, cela voudrait dire que les réseaux multimédias seraient contrôlés par les gouvernements... mais le problème reste entier : qui contrôlerait les gouvernements ? Les gouvernés voulant toujours être les gouvernants, comment résoudre ce problème ?

Un contrôle par le bas semble complémentaire, ce dernier supposant que les usagers souscrivent à un certain nombre de valeurs et de règles communes, une sorte de charte de l'information sur le Web.

La société de l'information semble bien remettre en cause les modes de gestion habituels du domaine politique : jusqu'à présent, l'exercice du pouvoir allait de pair avec le contrôle centralisé, exercé depuis le sommet de la pyramide vers sa base. La loi, la chose publique (la *res publica*) s'appliquent à des lieux précis ainsi qu'à des personnes bien identifiées. Dans la société en réseau, chaque nœud peut être délocalisé en fonction des besoins ou des usages. Par exemple, si un émetteur d'informations sur Internet est condamné pour des textes que la loi ou la morale réprovoque, le site émetteur pourra être délocalisé en un autre lieu, échappant ainsi à la juridiction locale. La jurisprudence appliquée à l'édition ou même à la télévision et à la radio n'est plus adaptée à la circulation de l'information sur les réseaux.

Il devient nécessaire de réfléchir à des formes nouvelles de contrôle et d'autorégulation faisant appel à la responsabilité de chacun.

Si l'on s'en remet aux usagers pour autoréguler la pratique et l'usage des sites proposés sur Internet, il est clair que la maturité des groupes sociaux n'est pas encore suffisante pour autoriser une régulation efficace. En revanche – comme le propose très justement Joël De Rosnay – si l'on fait reposer la responsabilité d'une telle régulation sur les familles, par exemple, on s'aperçoit que l'usage de simples logiciels peut

permettre de répondre aux problèmes de l'accès à des sites au contenu discutable. Ces logiciels sont paramétrables par les familles. Ensemble, parents et enfants peuvent décider du niveau de filtrage assuré par le logiciel.

IX.3 Comment construire une charte de bonnes pratiques de l'Internet ?

Si l'on décide d'élaborer une charte déontologique afin de réguler la production et la diffusion de contenus sur le Web, le problème de la notion même d'éthique va se poser.

L'éthique est-elle l'affaire de chacun ? Est-elle l'affaire de tous ?

Personnellement je ne vois pas l'éthique comme une donnée figée, je pense qu'elle doit, pour être légitime, être capable d'évoluer et de s'adapter à son environnement. L'éthique en temps de guerre ne saurait selon moi être la même que l'éthique en temps de paix. Mais pour beaucoup, l'éthique est une valeur dans l'absolu. Je suis d'accord sur le fait que la ligne directrice doit toujours être la même, mais je pense que dans l'application, l'adaptabilité est possible, voire souhaitable.

La question qui se pose alors est de savoir ce qui est bien dans l'absolu. Il semble qu'apporter une réponse à cette question, revienne à s'interroger sur le fondement même de la nature humaine. Quelle est cette base commune à tous les être humains, qui définit leur singularité et à la fois leur identité commune ?

Il est sans doute impossible de s'accorder sur tout ce qui est bien et sur tout ce qui est mal. Mais, même si l'on admettait qu'il n'existe qu'une dizaine de choses indiscutablement et universellement éthiques, ce serait à partir de ces convergences, qu'il faudrait construire une telle charte. Cette hypothèse a sûrement été émise lors des discussions qui ont préparé la rédaction de la Déclaration Universelle des Droits de l'Homme de 1948.

- **Le refus de la censure et de l'autocensure.**

D'autre part, même si l'on parvenait à se mettre d'accord sur une charte, il existe une difficulté supplémentaire : la censure.

Principe que tout journaliste est en droit de contester en vertu justement d'un article de la Déclaration des Droits de l'Homme, l'article 19 : « Tout individu a droit à la liberté d'opinion et d'expression, ce qui implique le droit de ne pas être inquiété pour ses opinions et celui de chercher, de recevoir et de répandre, sans considérations de frontières, les informations et les idées par quelque moyen d'information que ce soit ».

Ce refus d'autocensure par les journalistes eux-mêmes, est renforcé par les enjeux économiques de l'information de masse : le public est friand de révélations, d'images, de propos « scandaleux » qui sont officiellement et socialement réprouvés.

Il convient de remarquer que ces refus d'autocensure sont passibles de sanctions légales, lorsque l'on envisage les médias classiques (presse, télévision, radio) ; ces réflexions sont donc encore plus pertinentes dans le cadre du journalisme sur l'Internet, dans la mesure où le cadre légal y est beaucoup plus diffus et partiel, la réglementation étant particulièrement difficile à mettre en œuvre comme nous l'avons déjà souligné.

Pour les professions, on se réfère à la morale qui en sous-tend la pratique : c'est la déontologie. Mais la déontologie ne s'impose pas facilement. La déontologie ne

s'imposerait que s'il y avait des sanctions appliquées en cas de manquement, or c'est loin d'être le cas ; dans les médias, c'est la loi du profit qui souvent supplante l'éthique. En outre, les professionnels des médias s'opposent vivement à toute censure par les Etats, au nom de la liberté d'expression, condition de la démocratie.

Pour les professions, on se réfère à la morale qui en sous-tend la pratique : c'est la déontologie. Mais la déontologie ne s'impose pas facilement. La déontologie ne s'imposerait que s'il y avait des sanctions appliquées en cas de manquement, or c'est loin d'être le cas ; dans les médias, c'est la loi du profit qui souvent supplante l'éthique. En outre, les professionnels des médias s'opposent vivement à toute censure par les Etats, au nom de la liberté d'expression, condition de la démocratie.

L'éthique sur Internet doit donc avant tout relever du « bon sens », de la recherche de ce qui est bon pour l'Homme, et de la prohibition de tout comportement portant atteinte au respect et à la dignité de la personne humaine.

IX.4 D'importants enjeux sont liés au contrôle et à la diffusion de l'information.

En période de tension internationale et plus particulièrement de conflits armés locaux, les réseaux peuvent être utilisés comme une nouvelle arme d'information ou de désinformation. Déjà, certains pays qui se considèrent agressés, encombrant les serveurs de ceux avec lesquels ils sont en conflit par l'envoi de messages répétés créant des embouteillages sur les réseaux visés. D'autres émettent des virus qui ralentissent le fonctionnement des ordinateurs traitant des informations stratégiques. Une véritable guerre de l'information peut ainsi se dérouler sur Internet. Cette forme de conflit va s'accroître dans les années à venir. L'espionnage trouve également de nouvelles sources dans la société de l'information. Il est relativement facile de recueillir des données stratégiques circulant sur les réseaux. Des systèmes automatiques peuvent, à partir de mots-clefs, rechercher d'autres séquences de mots et même de phrases. Cette " *softwar* " ou guerre des réseaux représente le versant sombre de la société de l'information, mais il est inéluctable et va conduire à de nouveaux rapports de force.

Pour les pays en développement, l'avènement de la société en réseaux est à la fois une chance et un danger : une chance de pouvoir participer à l'économie mondiale grâce à des apports originaux dans différents domaines, une chance également de créer de la valeur ajoutée sans avoir à procéder à des investissements lourds dans les secteurs de l'économie traditionnelle, mais un danger d'isolement si les pays concernés ne savent pas se doter des moyens adaptés pour assurer leur présence à l'échelle mondiale.

Ils doivent pour cela surmonter une double difficulté : une contrainte intérieure d'abord, les gouvernements de la plupart des pays en développement craignant l'essor des moyens de communication individuels et permettant un traçage, tels les téléphones portables, les antennes paraboliques et évidemment Internet, moyens qui confèrent en effet aux individus un pouvoir démocratique susceptible de mettre en cause le pouvoir des régimes centralisés ; une contrainte de nature extérieure ensuite, représentée par l'ouverture au monde grâce aux autoroutes de l'information.

Aujourd'hui les débits des réseaux reliant les pays en développement au reste du monde sont très limités. Un cercle vicieux est en train de s'établir. Les investissements des grandes entreprises de télécommunication se réalisent aux endroits où les trafics sont importants. Il en résulte un accroissement des débits là où

ceux-ci sont déjà élevés. En revanche, des pays en nombre croissant sont de plus en plus isolés par des lignes à bas et moyens débits.

À l'intérieur du secteur des technologies de l'information, Internet joue un rôle de catalyseur et d'amplificateur. Les économistes reconnaissent que la société de l'information a un effet indirect sur la plupart des industries d'un pays et sur l'ensemble de l'économie. Création d'emplois, compétitivité industrielle et commerciale internationale, innovation, accélération de la circulation de l'information sont les résultats les plus probants de l'usage planifié des réseaux de communication dans tous les secteurs de la société.

Une entrée tardive dans la société de l'information représentera un risque de perte de compétitivité pour les industries et les pays qui n'auront pas su s'adapter à temps. C'est pourquoi il est nécessaire de s'inscrire le plus tôt possible dans ce que l'on appelle la " courbe d'apprentissage ". La forme caractéristique de cette courbe est celle d'une exponentielle. Sa première partie est plate et sa croissance faible. Puis elle croît pour se redresser presque à la verticale. Les investissements précédents produisent des intérêts réinvestis, conduisant à un effet d'amplification, même s'il s'agit d'investissements intellectuels ou immatériels. Ceux qui observent une telle courbe chez leurs concurrents et décident d'entrer dans le jeu en voyant les résultats obtenus, ou qui attendent l'ouverture de nouveaux marchés, risquent d'entrer trop tard dans la compétition. Il est indispensable en effet de passer par toutes les phases de la courbe d'apprentissage. Sans ces phases d'initiation, c'est à coup sûr l'échec par inadaptation à la croissance et à la progression des technologies ou des marchés.

L'impact des réseaux interactifs multimédia se fera sentir encore plus fortement dans les domaines de l'information, de l'éducation et de la culture. La logique actuelle d'accès à l'information est une logique de diffusion. Chaînes de télévision, de radio ou maisons d'édition émettent vers leurs téléspectateurs, auditeurs ou lecteurs de manière unilatérale, du haut de la pyramide vers sa base. La logique émergente résultant de la société de l'information est une logique de navigation. Loin d'être passifs à la base de la pyramide, récepteurs d'informations sans être capables d'échanges et d'interactions, les usagers deviennent actifs puisqu'ils doivent non seulement sélectionner les sources interactives d'informations, mais aussi entrer en contact avec d'autres usagers susceptibles de les aider.

Cette pratique modifie profondément les formes traditionnelles d'éducation. Plutôt que de se rendre en un lieu pour écouter un détenteur unique de savoirs, les internautes peuvent naviguer sur les réseaux à la recherche des informations qui leur permettront de construire leurs savoirs et leurs connaissances. Le risque est de brasser des informations en " surfant " d'un site à l'autre plutôt que d'acquérir des savoirs, d'où l'importance d'une formation dès l'école en vue de favoriser l'intégration culturelle car le Web permet, si l'on sait l'utiliser d'accéder à une richesse culturelle sans frontières.

Le nouveau rôle des enseignants serait alors d'aider à faire émerger l'intelligence collective de ses élèves, de ses étudiants, de les aider à s'orienter sur le chemin électronique de la connaissance.

Internet constitue un nouveau système de communication interactif permettant d'intégrer des données dans des informations, des informations dans des savoirs, des savoirs dans des connaissances et des connaissances dans des cultures. Cette intégration progressive est représentative de l'acquisition d'une culture moderne. Elle se distingue de la simple culture des cultivés " qui savent des petits riens sur tout " ou

de celle des spécialistes qui " savent tout sur des petits riens ". La société de l'information suscite la réflexion sur un plan ou un projet capable d'inclure les éléments technologiques amplifiant les modes d'apprentissages traditionnels.

« On sait que contrairement aux bibliothèques et aux CDI, le Web est une jungle, un océan, un fouillis ou une poubelle, selon l'appréciation. Ce qui est patent, et qui constitue d'ailleurs l'un des enjeux éducatifs les plus forts, c'est bien ce retournement de la validation de l'information : jusqu'alors effectuée "en amont" de la chaîne de production de l'information, d'abord par les chercheurs et les auteurs, qui n'écrivent pas (théoriquement) n'importe quoi, puis par les éditeurs, qui ne publient pas tout ce qui s'écrit, ensuite par les libraires, qui ne vendent pas tout ce qui se publie et enfin par les bibliothécaires-documentalistes, qui n'achètent pas tout ce qui se vend, la validation de l'information (terme générique sous lequel on mettra l'évaluation, la sélection, le filtrage...) s'opérait à différents niveaux, par différents acteurs, selon différentes modalités et pour différentes finalités. Ce schéma, toujours valable dans le monde "traditionnel" de l'édition ou de la production scientifique, n'est plus celui du Web : la validation de l'information est ici généralement reportée sur l'utilisateur, "en aval", avec tous les problèmes, les risques et les dégâts possibles¹⁴² ».

¹⁴² Lettre de l'URFIST n°34 www.urfist.cict.fr/lettres/lettre34/lettre34-31.html

Conclusion et perspectives

...Et que même les plus belles choses ont une fin...

X. Conclusion et perspectives

Il existe de bonnes sources d'informations riches et puissantes sur Internet, mais elles sont payantes, et de ce fait trop souvent boudées pour des sources gratuites mais moins fiables ou moins directement exploitables.

Ces sources gratuites sont accessibles via les outils de recherche grand public, qu'il s'agisse de moteurs de recherche, d'annuaires thématiques ou de métamoteurs.

Ces outils ont réussi à s'imposer dans le cœur des internautes et constituent de véritables portails vers l'information.

Or, la gratuité d'accès à l'information qu'offrent ces outils à l'internaute n'est qu'apparente. En effet, les sociétés de services d'accès à l'information les plus utilisées, qu'il s'agisse de Google ou de Yahoo!, sont avant tout des entités commerciales, qui, pour exister et se développer, ont dû dépasser les modèles économiques traditionnels – telles que les bandeaux publicitaires – pour mettre en avant de nouveaux modèles économiques, plus subtils, plus insidieux, comme le positionnement payant, l'insertion payante ou encore le coût par clic, qui ne sont pas toujours signifiés aux utilisateurs de la manière la plus transparente qui soit.

Ce manque de transparence et, plus généralement, l'ingérence du commerce dans le monde de la recherche d'information a plusieurs conséquences, notamment celle d'introduire un coût pour l'internaute. Un coût en terme de temps passé à trier l'information commerciale de l'information non commerciale, et les résultats naturels des résultats dus à des accords commerciaux entre les moteurs et les annonceurs.

D'autre part, ce manque de transparence défavorise les internautes non avertis, c'est-à-dire les internautes n'ayant pas reçu de formation en matière de recherche de contenus. En effet, pour tirer bénéfice d'une recherche d'information, il est nécessaire de connaître à la fois le fonctionnement des moteurs de recherche du point de vue de leur architecture logicielle, mais aussi les modèles d'affaires régissant l'économie des moteurs, et enfin, de suivre une méthodologie réfléchie, ordonnée, permettant d'épurer progressivement l'information récoltée pour dissocier le vrai du faux.

Une solution semble venir du recoupement d'informations, ce qui implique d'avoir clairement identifié sa problématique et son sujet de recherche, puis de rechercher de l'information dans les sources validées en amont et dans celles qui doivent être validées a posteriori (absence de date, d'auteur, de contexte clairement défini).

D'autre part, d'autres types de recherche d'informations préexistaient au Web et aux moteurs de recherche, et ils ne doivent surtout pas être oubliés : la recherche par les documentalistes, en centres ou en bibliothèques demeure un canal d'information capital, et interactif puisqu'il est toujours possible d'échanger avec un humain ! L'internaute doit intégrer la logique de passerelles entre les médias et rechercher l'information sur le Web et hors Web, car exploiter et valoriser l'information c'est être capable de recouper des informations, des outils de recherche et des sources d'informations différentes.

Quelle que soient les sources choisies, l'internaute devra toujours en référer à son bon sens et faire preuve d'esprit critique par rapport au monde qui l'entoure.

De même, le fait de rechercher de l'information sur des sites dédiés à l'information scientifique et technique semble plus sûr, mais le risque est alors que ces informations ne soient contrôlées en amont par des gouvernements, qui n'autoriseraient que la diffusion des informations ne mettant pas en péril le fonctionnement de la nation.

En effet, si à l'ère « industrielle », ceux qui maîtrisaient les moyens de production d'énergie possédaient le pouvoir, aujourd'hui, c'est l'information qui constitue une arme stratégique.

Dans leur recherche du contrôle, les grandes organisations, et les nations ont bien compris que la puissance appartient à celui qui maîtrise les flux informationnels et les contenus qui s'y trouvent, les télécommunications et les medias représentent des secteurs stratégiques, objets d'une guerre économique terrible sous une forme apparemment anodine. Maîtriser les nouveaux types d'attaques informationnelles introduits par le détournement des modèles d'affaires des outils de recherche (rumeurs, désinformation...) est aujourd'hui une nécessité, et toute entité, commerciale, ou politique, ne disposant pas d'une cellule d'intelligence économique, pilotée par des experts en recherche d'information et en veille économique, prend un risque qui peut lui coûter très cher, d'autant qu'il est de plus en plus simple de diffuser de l'information sur Internet (blogs, etc.).

Extension du monde réel, le Web est donc un monde virtuel au sein duquel chaque individu, chaque Etat, chaque Nation, peut exister et s'affirmer.

Et s'il est fort probable que la logique d'intégration du Web l'amènera à voir coexister toutes les formes de médias (qu'il s'agisse de vidéos, musiques, images, sons, etc.), il restera dépendant du texte.

Fort de ces deux constats, il est une composante de l'identité culturelle d'une Nation qui doit attirer notre attention : la langue.

La langue est l'expression de l'identité d'une Nation, or, malgré la vocation du Web à promouvoir le multilinguisme, de nombreuses craintes s'élèvent régulièrement contre une tendance à l'homogénéisation des langues vers l'anglais. Les langues des internautes ne sont pas menacées, mais paradoxalement, le multilinguisme semble l'être : il existe toujours des contenus rédigés par des internautes dans leur langues maternelle, mais le réflexe de construire des sites multilingues n'est pas encore pavlovien chez les administrateurs de sites Web. Or certaines informations deviennent inaccessibles pour l'internaute simplement parce qu'il ne maîtrise pas la langue dans laquelle ces informations sont rédigées.

D'autre part, si l'on focalise sur l'information que l'on cherche comme un prédateur, il ne faut pas oublier qu'en tant qu'internaute consommateur, on constitue également une proie sur le net, une cible pour les systèmes d'intelligence marketing des entreprises qui surveillent nos habitudes et nos comportements de consommateurs à l'aide des traces que nous laissons souvent sans le vouloir (ou sans le savoir), comme le journaliste Shi Tao, qui, comble de la profession, s'est vu dénoncé aux autorités par un moteur de recherche d'information.

Tout est fait pour plaire aux internautes et les outils évoluent, se personnalisent. Les systèmes adoptent une logique d'intégration où les différents supports d'information se croisent, s'imbriquent et dressent des passerelles les uns entre les autres (les publicités télévisuelles renvoient de plus en plus souvent les consommateurs sur un site Web).

Demain, c'est la tendance du nomadisme qui se renforcera et l'information deviendra portable, en mouvement (Blackberry, téléphonie 4G...). Les medias traditionnels critiquables sur leur mode unidirectionnel de diffusion de l'information, devront rapidement évoluer et migrer sur le canal Internet pour profiter du feedback des consommateurs, hier passifs, aujourd'hui acteurs et bien informés... le prochain challenge de la société de l'information de demain, et qui est une question d'actualité aujourd'hui, sera de définir les modes de contrôle d'Internet, cet outil puissant capable

de toucher un grand nombre de personnes, de façon ciblée en très peu de temps en faisant fi des frontières physiques, et qui a le pouvoir de se retourner contre ceux qui jusqu'alors détenait le monopole des médias : les gouvernements.

L'enjeu de la recherche d'information de demain sera de conquérir de nouveaux marchés mobiles (m-commerce) s'inscrivant dans cette logique constante d'intégration des technologies entre elles et de nomadisme, comme par exemple la recherche par sms (envoi d'un sms à Yahoo ! par exemple, puis accès à un portail de services). Autre tendance de l'industrie, dépasser la recherche par motif (chaîne de caractères) pour accéder à une recherche sur les concepts, et passer de la gestion documentaire, à la gestion des connaissances.

Enfin, l'accroissement des débits offrira aux entreprises et aux particuliers de nouveaux modes de diffusion de l'information avec l'introduction de la télévision interactive (Webcasting), de graphismes professionnels en trois dimensions (utilisables pour la chirurgie, l'architecture, l'ingénierie, la vente de produits...) et du Triple Play, qui combine la télévision, la voix et l'Internet, et qui nous permettra bientôt d'accéder à l'Internet sur nos écrans de télévision.

Il est à parier que le prochain outil phare du Web qui recevra, à l'instar du moteur de recherche, les faveurs de la communauté Internet, ne répondra non plus uniquement à un besoin en recherche d'information, mais également à un besoin en traitement automatique de ces informations pour dépasser la reconnaissance et l'extraction de mots-clés et extraire directement les concepts les plus pertinents... pour un Web sémantique...

Bibliographie

- Bibliographie

BERGMAN (2001), « The Deep Web : Surfacing Hidden Value », BrightPlanet.com LLC 2001.

BROUSSEAU Eric (2001) « Commerce électronique : ce que disent les chiffres et ce qu'il faudrait savoir », ÉCONOMIE ET STATISTIQUE N° 339-340, 2000 - 9/10.2001. Disponible sur http://www.insee.fr/fr/ffc/docs_ffc/ES339-340F.pdf

CHARTRON Ghislaine. « De l'information spécialisée à l'information élaborée : problèmes de modélisation », 8e congrès SFSIC, Lille, 21-22 Mai 1992.- Lille : CREDO/ SFSIC/ Université Charles de Gaulle, Lille 3, 1992.- 462p. Meyriat et Estival, 1981, p.84)

CHAUMIER Jacques, DEJEAN Martine, « L'indexation documentaire, de l'analyse conceptuelle à l'analyse morphosyntaxique », Documentaliste, vol.27, n°6, novembre-décembre 1990, pp.275-279

DE LOUPY Claude (2000). « Multilinguisme et document numérique : la dimension technique à l'épreuve du codage des caractères ». © "Solaris", n° 6, Décembre 1999 / Janvier 2000. Disponible sur <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d06/6loupy.html>

DE ROSNAY, Joël (1999), « La société de l'information au XXIe siècle », article paru dans le numéro de Ramsès 2000 de l'Institut Français des Relations Internationales (Ifri). 1999. Disponible sur http://csiweb2.cite-sciences.fr/derosnay/articles/Ramses_FinalPrint.html

ESTIVAL, Robert – MEYRIAT Jean (1981).- "La dialectique de l'écrit et du document. Un effort de synthèse".- in : *Schéma et schématisation*, n°14, 2e trim. 1981.- p.82-91

FAYET-SCRIBE, S. (1997) « Chronologie des supports, des dispositifs spatiaux, des outils de repérage de l'information », © "Solaris", n° 4, Décembre 1997. Disponible sur http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d04/4fayet_3bib.html

GULLI & SIGNORINI (2005), "The indexable web is more than 11.5 billion pages", Università di Pisa, Département d'Informatique. 2005. disponible sur: www.cs.uiowa.edu/~asignori/web-size/size-indexable-web.pdf

LEBERT Marie (1999) « Le multilinguisme sur le web » février 1999. Disponible sur <http://www.cefrio.qc.ca/projets/Documents/multifr0.htm>

LE COADIC Y F (1998) .*Le besoin d'information*. Adbs édition, 1998.

LUCAS André, DEVEZE Jean, FRAYSSINET Jean, « Droit de l'informatique et de l'Internet », PUF, novembre 2001, n° 704.

MUCCHIELLI (1995) . *Les Sciences de l'Information et de la Communication*, éd. Hachette, 1995

RAKOVEC Pablo (2002), « Les mécanismes de recherche d'informations utilisés par Google ».CNAM, Languedoc-Roussillon. Disponible sur http://cj.tronquet.free.fr/cnam/pr_detail.php?dl=54

SALGADO LEMOS Daniel (2003) « Technologies de l'information et de la communication : coordination des réseaux pour le développement durable des zones enclavées. Le cas de la Békaa, Liban » Thèse de Master of Sciences du CIHEAM-n°59, p.12.2003. Disponible sur : <http://ressources.iamm.fr/theses/59.pdf>

SARACEVIC, T. (1996). *Relevance reconsidered*. In Ingwersen P, eds. *Colis2*; 1996 October 13- 16; Copenhagen, Danmark; PP 201-218

SCHAMBER, L. (1994). *Relevance and information behavior*. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology*. 29 (pp. 3-48). Medford, NJ: Learned Information

Articles/Rapports/Enquêtes/ Études :

BARGE, S. (2005) « Google, le nouveau maître du monde » *Capital* octobre 2005, p.36

« Google encore attaqué sur de la fraude au clic » Reuters jeudi 30 juin 2005, repris par <http://www.lexpansion.com/html/A133169.html>

CAQUELIN P-H, GATCAN D et VETILLART C (1999-2000) Mémoire, IEP de Paris « De Monica à DSK : quelle éthique pour le journalisme sur l'Internet ? » dans le cadre de la conférence TIC « Internet et le numérique - Enjeux et usages » disponible sur <http://ethique.ifrance.com/hypermarche.html> 1999-2000.

COURIER Serge (2005). « On trouve tout dans les encyclopédies en ligne ». *Management*, Avril 2005 pp.110-113.

DONNEDIEU DE VABRES Renaud (2004), Discours donné à l'occasion de la réception en l'honneur de la Fédération Européenne des institutions linguistiques nationales, lundi 8 novembre 2004, disponible sur <http://www.culture.gouv.fr/culture/actualites/>

HASKI Pierre (2005) « Comment Yahoo a sacrifié un dissident chinois à ses intérêts » *Libération*, 16 septembre 2005. Disponible sur <http://www.liberation.com/imprimer.php?Article=324069>

PLUYETTE C., et al. (2005) « Où et comment trouver les bonnes infos sur Internet » *Management*, N° 125, novembre 2005, pp. 34-58
« Les moteurs de recherche relayent les messages publicitaires ». *Le Figaro* du 11/07/2005 d'après une étude UDA/Overture/Isobar

SERRES Alexandre (2004) « Recherche d'information sur Internet : où en sommes-nous, où allons-nous ? », disponible sur <http://savoirscdi.cndp.fr/CulturePro/actualisation/Serres/Serres.htm>

SERRES Alexandre (2004) « Les sept grandes tendances de la recherche d'information sur Internet » Lettre de l'URFIST n°34 Disponible sur www.urfist.cict.fr/lettres/lettre34/lettre34-31.html

Contrôle d'internet: accord à Tunis entre les Etats-Unis et les autres pays, dépêche **AFP**, 15/11/2005, disponible sur http://www.lintelligent.com/gabarits/articleDEP_online.asp?art_cle=AFP72605contrsyapse0

RAPPORTS

Rapport Détaillé, "L'évolution de la presse écrite dans la perspective des nouvelles technologies multimédias" en 1994; cité sur <http://www.men.minefi.gouv.fr/webmen/revuedeweb/presseligne.html>

Rapport Bourdier "La presse et le multimédia" en 1997), disponible sur <http://www.ladocumentationfrancaise.fr/rapports-publics/974055500/index.shtml>

Rapport Bangemann intitulé «Europe and the global information society », disponible sur: <http://europa.eu.int/ISPO/infosoc/backg/bangeman.html>

Rapport final concernant le processus de consultations de l'OMPI sur les noms de domaine de l'Internet, 30 avril 1999, <http://wipo2.wipo.int>

Danny Sullivan, *Europe Paid Placement Warriors*, The Search Engine Rapport, 4 février 2002. disponible sur <http://searchenginewatch.com/sereport/article.php/2164541>

ÉTUDES

NUA, <http://www.nua.com/surveys>

Carat Expert (2005) Paris, le 30 juin 2005. L'Union des annonceurs (UDA), Overture France et Isobar. « Les moteurs de recherche, relais des médias traditionnels ». Disponible sur www.uda.fr/html/infos/actualite/overture/CP_ETUDE_FINAL.pdf et www.overturenbrief.fr/div/etude_juin2005/UDA-Overture-Isobar.pdf

Livre Blanc Influo (2002) « Moteurs de recherche, où est la technologie ? » disponible sur http://h2ptm.univ-paris8.fr/je2002-2003/20030114_Recherche_et_fouille_information_sur_le_Web/prez/Influo%20WP%20moteurs%20V1.pdf

Livre Blanc Benchmark Group/Netbooster (2005) "Les enjeux du référencement de sites internet" disponible sur http://www.journaldunet.com/livres_blancs/netbooster/netbooster.pdf

Nielsen/NetRatings (2003) « Le classement mondial » disponible sur http://www.journaldunet.com/tops/top_worldratings.shtml ou <http://www.nielsen-netratings.com/>

OneStat (Juin 2002 à Mai 2004) « baromètre mondial des outils de recherche » disponible sur <http://www.indicateur.com/barometre/>

Leslie Marable (2003) “False Oracles: Consumer Reaction to Learning the Truth About How Search Engines Work” chercheur/écrivain, pour Consumer Reports WebWatch en juin 2003. disponible sur <http://www.consumerwebwatch.org/dynamic/search-report-false-oracles.cfm>

HEITZMANN Raymond, DAYAN Martine (2004)« Tableau de bord du commerce électronique ». Mission pour l'économie numérique. Baromètre Multimédia de Médiamétrie, CSA, CSA Opinion ,TMO, Médiangles. Service des Etudes et des Statistiques Industrielles – 2004. disponible sur <http://www.men.minefi.gouv.fr/webmen/themes/eco/tbce91204.pdf>

Computer Industry Almanac – Market Research Reports (<http://www.c-i-a.com/>)

STATISTIQUES DE L'INTERNET

Répartition des internautes par langues
<http://www.internetworldstats.com/stats7.htm> , 21 novembre 2005

Répartition de la population internaute
<http://www.internetworldstats.com/stats.htm> , 09 novembre 2005

Webographie

Encyclopédies

www.wikipedia.org
www.webencyclo.com
<http://fr.wictionary.org>
www.imdb.com
www.discogs.com
www.gracenote.com/music
www.aquabase.org
www.wikif1.org
www.wikitavel.org

Autres sites Web cités :

Site d'André Santini : <http://www.andre-santini.net/>
Site de Roland Aberlin sur la lune : <http://r.aberlin.free.fr/lune/lune.htm>
Liste d'outils utilisés en annexes, disponible sur: <http://www-Rocq.INRIA.Fr/>
Le journal du Net
<http://www.acadie.net/chronique/contenu.cfm?identification=6977>
<http://www.silicon.fr/getarticle.asp?ID=11953> (75 millions de sites Web, étude de Netcraft)

<http://www.indicateur.com>

Outils de recherche Web Invisible

<http://www.profusion.com>
<http://www.invisible-web.net>
<http://www.completeplanet.com>
<http://urfist.uni-lyon1.fr>
<http://www.ccsd.cnrs.fr>
<http://arxiv.org>
<http://www.sofymetal.com>
<http://www.archive.org> .

Métamoteurs/ Moteurs

<http://www.ariane6.com/>
<http://www.metacrawler.com/>
<http://www.kartoo.com/>
<http://vivisimo.com/>
<http://www.ez2find.com/>
<http://www.findforward.com/>
<http://www.eu.ixquick.com/>
<http://www.copernic.com/meta/web/>
<http://www.killerinfo.com/>
<http://www.surf wax.com/>
<http://www.grokker.com/>
<http://www.dogpile.com>
www.exalead.com/search
www.marumushi.com/apps/newsmap
http://www.droitdunet.fr/par_profils/lecture.phtml?type=profil_consommateur&it=5&id=4
<http://www.google.com/psearch>
<http://myweb2.search.yahoo.com/>
<http://www.eurekster.com/>

<http://www.kumc.edu/service/dykes/classes/search.html>
<http://www.uhb.fr/urfist/Supports/FormaDoctor/FormDoctMethodo.htm#RessourcesA>
[utoformation](#)
www.technorati.com
www.blogpulse.com
www.blogsearch.google.com
www.rezoweb.com
www.yahoo.groups.com
www.groups.google.com
www.enfin.com
www.lesannuaires.com
www.indicateur.com. »¹⁴³

SITES DEDIES AUX MOTEURS DE RECHERCHE

“The History of Yahoo ! How it all started...”, Disponible sur
<http://docs.yahoo.com/info/misc/history.html>
www.abondance.com
www.SearchEngineWatch.com
www.webrankinfo.com
www.secrets2moteurs.com

DROIT

<http://www.juriscom.net>
SARL Distrimart c/ SA Safi, CA Paris, 14 mars 2001, disponible sur *Juriscom.net*,
<http://www.juriscom.net/txt/jurisfr/ndm/caparis20010314.htm>

Société Pierre Fabre Dermo Cosmétique et autres c/ Alain B, CA Versailles, 2 décembre 1999, disponible sur *Juriscom.net*
<http://www.juriscom.net/txt/jurisfr/ce/caversailles19991202.htm>

COHEN Murielle, Avocat-online, « Conseils juridiques de Murielle Cahen »
Disponible sur <http://www.murielle-cahen.com/>

DIMEGLIO, Arnaud (2001), *Les contrats de référencement dans l’Internet*”,
Comm.Com.Elect., mars 2001, p.14.

LETOURNEAU Emmanuelle (2004) “Le contrôle de l’information sur Internet”
<http://www.barreau.qc.ca/journal/frameset.asp?article=/journal/vol33/no20/surlenet.html>

Voir également http://www.rsf.org/rubrique.php3?id_rubrique=432 (internet sous surveillance 2004)

Loi n° 93122 du 29 janvier 1993 relative à la prévention de la corruption et à la transparence de la vie économique et des procédures publiques, *JO* du 30 janvier 1993.

¹⁴³ Capital N°125, Novembre 2005

SA Citycom c. SA Chanel, CA Paris, 3 mars 2000, disponible sur [Foruminternet.org](http://www.foruminternet.org/documents/jurisprudence/lire.phtml?id=79),
<http://www.foruminternet.org/documents/jurisprudence/lire.phtml?id=79>

SA Citycom c/ SA Chanel CA Paris, 3 mars 2000, disponible sur [Foruminternet.org](http://www.foruminternet.org/documents/jurisprudence/lire.phtml?id=79)
<http://www.foruminternet.org/documents/jurisprudence/lire.phtml?id=79>

TGI Paris, 29 octobre 2002, décision disponible sur le site www.legalis.net/

United States District Court, Central District of California, *Nissan Motor Co. Ltd v. Nissan Computer Corporation*, 29 août 2001
<http://searchenginewatch.com/sereport/02/02-bodysolutions.html>

CJCE, 4 novembre 1997, *Dior c. Evora*, PIBD 1999 n°676, III, p.221.
CA Rennes, réf., 30 mars 2000, JCP ed.E 2000, n°48, p.1902
La gestion des noms et adresses de l'Internet : Questions de propriété intellectuelle”,
Laurent Bailliard, “ *Quand le référencement flirte avec l'illégalité* ”, VnUNET.fr 3
janvier 2002
“ *Le Position squatting, nouveau défi des marques sur Internet* ”, Cvfm.com,
décembre 2001.

CA Paris, 22 février 1995, D.S., 1996, Somm.250

TGI Paris, réf., 31 juillet 2000, *M. Bernard D. c/ Sté AltaVista Company*, disponible
sur www.juriscom.net/txt/jurisfr/cti/tgiparis20000731.htm

Landericht Hamburg, 16 février 2000, *Estée Lauder c. Fragrance Counter*
<http://searchenginewatch.com/sereport/02/02-bodysolutions.html>

Art. 121-7 al.1 Nouveau Code Pénal

« La Protection de la Propriété Intellectuelle face aux Nouvelles Technologies de
l'information et de la Communication »
http://memoireonline.free.fr/propriete_intellectuelle.htm

FAUCHE, Yann (2000) “Anonyme sur Internet disponible sur le web invisible via la
Waybackmachine :
http://web.archive.org/web/20010405174525/yann.fauche.free.fr/anonyme_net.php3

L'ICANN : <http://www.icann.org/tr/french.html>

LA CNIL - Commission nationale de l'informatique et des libertés : www.cnil.fr/

METHODOLOGIE & RECHERCHE D'INFORMATION

http://www.accart.nom.fr/Cours/RechInfo_01.html

Méthodologie de la recherche documentaire

<http://users.11vm-serv.net/aeris/cours/chercher/methodes.html>

ANNEXES

En bonus : une boîte à outils complémentaire pour le chercheur d'information sur le web

- ANNEXES

TRADUCTEURS AUTOMATIQUES (LISTE EXHAUSTIVE SUR <http://www-rocq.inria.fr/>)

[SYSTRAN](#)

Traduction d'un texte ou d'une page Web entre anglais ou français d'une part et allemand, anglais, espagnol, français, italien ou portugais d'autre part.

[AltaVista's Translation Assistant](#)

Traduction d'un texte ou d'une page Web entre allemand, anglais et français et entre anglais et espagnol, anglais et italien, anglais et portugais, avec une ouverture vers le russe. Utilise la technologie SYSTRAN.

[Reverso](#)

Traduction d'un texte ou d'une page Web entre allemand, anglais et français, avec une ouverture vers l'espagnol et le russe. Utilise la technologie Softissimo.

[PROMT-Reverso Online](#)

Traduction d'un texte ou d'une page Web entre anglais, français et russe, avec une ouverture vers l'allemand et l'espagnol. Utilise la technologie Systran.

[VoiLa traduction](#)

Traduction d'un texte ou d'une page Web entre allemand, anglais, français et russe, avec une ouverture vers l'espagnol et l'italien. Utilise la technologie Softissimo.

[FreeTranslation de Transparent Language](#)

Traduction d'un texte ou d'une page Web entre anglais et allemand, anglais et espagnol, anglais et français, avec une ouverture vers l'italien, le norvégien et le portugais.

ENCYCLOPEDIES

[InfoPlease](#)

Encyclopédie américaine très complète accompagnée d'un dictionnaire, d'un almanach, de dépêches d'actualité thématiques, etc.

[L'Encyclopédie Hachette Multimédia](#)

Encyclopédie francophone des Éditions Hachette.

[L'Encyclopédie Microsoft](#)

Encyclopédie francophone de Microsoft accompagnée d'un dictionnaire bilingue. Produit d'appel pour suggérer l'achat de la version sur cédérom, pas très pratique mais à consulter néanmoins.

[Encyclopedia.com de Electric Library](#)

Encyclopédie américaine. Ne sait pas faire la différence entre le gouvernement, l'État et la Nation...

[L'Encyclopédie canadienne de la fondation HISTOR!CA](#)

Encyclopédie en français et en anglais. Intéressante, malgré une navigation peu aisée et un contenu très centré sur le Canada.

DICTIONNAIRES

[Free On-Line Dictionary Of Computing \(FOLDOC\)](#)

14169 termes anglais en relation avec les ordinateurs, le logiciel et les réseaux, édités par l'Imperial College.

[ARTFL Project : French-English Dictionary Form](#)

[English-French dictionary of common computing terms de l'université Reine Mary à Londres](#)

[EuroDicAutom WWW Interface](#)

La banque de données terminologiques multilingue du service de traduction de la commission européenne. Disponible en anglais.

[The LOGOS Dictionary \(FR\)](#)

Bel effort de traducteurs professionnels pour réaliser une banque de données multilingue.

[Merriam-Webster's WWWebster Dictionary](#)

[Hypertext Webster Gateway](#)

[Recherche détaillée de documents de l'Union Internationale des Télécommunications](#)

[ARTFL Project : Webster's Revised Unabridged Dictionary, 1913 Edition](#)

[Wordsmyth Educational Dictionary-Thesaurus](#)

Dictionnaire de l'Anglais US moderne.

[WordNet: A Lexical Database for English](#)

[Dictionnaire des synonymes de l'université de Caen](#)

Chaque entrée est suivie d'une liste de synonymes, qui ne distingue ni les catégories ni les nuances de sens, et un traitement informatique permet l'obtention des différences de sens (sens élémentaires du mot-vedette) par une analyse mathématique des relations synonymiques de cette liste.

[OneLook Dictionaries](#)

Accès direct à plus de 2 millions de mots indexés par OneLook en plus de 5 langues.

[The IT-specific encyclopedia \(whatis?com\)](#)

Des milliers de termes de l'informatique, en américain.

[Answer.com](#)

[Recherche dans le *Collins Large English Dictionary \(unabridged\)*](#)

Excellent dictionnaire anglais.

[Recherche dans le *Collins French Dictionary*](#)

Un bon dictionnaire bilingue anglais et français.

[Trésor de la Langue Française Informatisé \(TLFi\)](#)

Le dictionnaire des XIXe et XXe siècles en 16 volumes et 1 supplément, élaboré par l'INALF de 1971 à 1994. Une superbe référence, dont l'interface à l'ergonomie médiocre offre des possibilités de consultation époustouflantes. Méfiance toutefois en cas de recherche infructueuse.

GLOSSAIRES

[Glossaire typographique et linguistique \(Babel\)](#)

Réalisé par Alis Technologies pour l'ISOC lors d'une étude sur la mondialisation d'Internet. Une référence précieuse.

[L'ABC des réseaux](#)

Publié par la DGLF, le glossaire de Jean-Alain Hernandez et René Joly (E.N.S.T.). Commence à dater (1994, apparemment).

[Glossaire informatique des termes de la CMTI.](#)

Les termes publiés ou en voie de publication (en 1997) de la commission ministérielle (française) de terminologie de l'informatique. Périmé, n'a plus qu'un intérêt « historique » pour les lexicographes.

[Dictionnaire de l'internaute \(IBM Canada\)](#)

Plus de 2500 entrées et 1800 définitions françaises dans les domaines suivants : commerce électronique, infographie, informatique, internet, micro-informatique, multimédia, réseautique, sécurité informatique, technologies de l'information, télécommunications.

[WinFiles.com Internet Hyper-Glossary](#)

Très bon petit glossaire technique à l'usage du commun des mortels anglophone. Malheureusement stable depuis 1996 (l'ADSL n'est pas mentionné, par exemple).

[Glossaire de l'Internet de Compuform](#)

[Glossaire de termes relatifs à Internet](#)

Publié par la DGLF depuis 1996, le glossaire de Jean-Karim Benzineb, traducteur au Conseil de l'Europe.

[Le Jargon Français](#)

[TechEncyclopedia \(CMPnet\)](#)

[WebMonkey Glossary](#)

[The Memory Management Glossary](#)

[Glossaire du Service ATO de l'UQAM](#)

Glossaire de vulgarisation pour le domaine du traitement automatique de la langue en général et de l'analyse de texte par ordinateur en particulier.

[RÉTIF](#)

Concerne les termes de l'informatique et des réseaux.

[Encart du N°14 du 8 avril 1999 du Bulletin Officiel de l'Education Nationale](#)

Ce sont eux qui l'appellent comme ça... Reprend la liste des termes, expressions et définitions adoptés de décembre 1997 à mars 1999 par la Commission générale de terminologie et de néologie, dans les domaines de l'architecture, des carburants, du courrier électronique, de l'économie et des finances, de l'informatique, de l'Internet, du sport, des transports, de l'unité de monnaie européenne.

[Listes de termes publiés au Journal officiel*](#)

Présentés par la DGLF, voici la liste des listes de termes, expressions et définitions adoptés et publiés au Journal officiel de la République française sous l'égide de la Commission générale de terminologie et de néologie (en application du décret N°96602 du 3 juillet 1996 relatif à l'enrichissement de la langue française).

[Base de données CRITER](#)

La base de données terminologiques du dispositif d'enrichissement de la langue française, constituée à partir des listes publiées par la Commission générale de terminologie et de néologie. Un peu fruste et pas toujours à jour, mais cela a le mérite d'exister.

[Le Signet](#)

6000 fiches terminologiques bilingues dans le secteur des technologies de l'information, proposées par l'Office de la langue française du Québec.

[Base de connaissances en Télécommunication et Informatique](#)

Intéressant glossaire technique bilingue anglais-français.

[Dictionnaire panlatin des termes de base de l'informatique du CRETAL*](#)

Un travail réalisé entre 1995 et 1997 à l'Université Paris III « Sorbonne nouvelle », en catalan, espagnol (Espagne, Argentine, Cuba et Mexique), français (Québec et France), italien, portugais (Portugal et Brésil) et roumain --- auxquels a été ajouté l'anglais.

[Acrodict, dictionnaire francophone des acronymes informatiques](#)

Glossaire revendiquant 4000 acronymes et sigles.

[Sun Global Glossary](#) _

Plus de 2000 termes autour de la technologie et des produits de Sun. Tout est abordé, le matériel, la réseautique, l'environnement graphique, le système d'exploitation (Solaris, bien entendu), le langage Java...

LEXIQUES

[Lexique des néologismes Internet](#)

Rassemblés par Guy Brand et Jean-Pierre Kuypers, des termes communément utilisés par des traducteurs et adaptateurs francophones de logiciels.

[Lexique de dépannage orthographique](#)

Donne l'orthographe exacte à partir d'une graphie approximative.

[The WorldWideWeb Acronym and Abbreviation Server](#)

Près de 18 000 acronymes de tous domaines. Très complet.

[The canonical abbreviation/acronym list](#)

[Vocabula computatralia](#)

[ISI glossary of statistical terms](#)

Glossaire de statistique compilé par le INTERNATIONAL STATISTICAL INSTITUTE. Très complet, apparemment.

LISTES DE DICTIONNAIRES, GLOSSAIRES ET LEXIQUES

[Dictionnaires et Lexiques](#)

Recense les dictionnaires francophones en ligne ainsi que des lexiques classés par catégories : dictionnaires multilingues, dictionnaires de langue française, grammaire, morphologie, orthographe, *et cetera*.

[Dictionnaires \(Leximagne\)](#)

[Métalexis](#)

[Lexique du CAFI sur l'informatique et les domaines connexes](#)

[Your Dictionary](#)

Un portail linguistique qui répertorie plus de 1500 dictionnaires en ligne dans 230 langues. Nombreuses rubriques : vocabulaire, grammaire, thésaurus, *et cetera*.

Succède à « A Web of On-line Dictionaries ».

[OneLook Dictionaries](#)

Les 600 et quelques dictionnaires en ligne indexés par OneLook, rangés par catégorie.

[Sites de terminologie proposés par le service linguistique de l'INRA](#)

Principalement dans des domaines scientifiques.

[Liste proposée par le Service des bibliothèques de l'Université de Sherbrooke, au Canada](#)

Dictionnaires de langues, dictionnaires spécialisés en anglais et en français, encyclopédies...

[Listes de dictionnaires proposées par les services linguistiques centraux de l'administration fédérale suisse](#)

[Ressources remarquées par le traducteur John Lejderman](#)

La page est intitulée *Dictionnaires et glossaires anglais-français sur l'informatique, les systèmes d'information et les technologies de l'information*

AUTRES REFERENCES A CONSULTER

[Échantillon analytique du dictionnaire de l'Académie française](#)

Vitrine du projet InfoDAF d'informatisation des huit premières éditions du dictionnaire de l'Académie française.

[Syntax Student's Companion](#)

Appliquette d'édition visuelle tout en anglais d'arbres syntaxiques, à partir de ressources encodées en XML.

[The Alternative French Dictionary \(argot\)](#)

[Liens argotiques du site « ABC de la langue française »](#)

[The Dictionary of the Heart](#)

Vous voulez offrir un bouquet d'oeillets et de roses jaunes ? Apprenez le langage des fleurs pour les utiliser à bon escient !

[Ressources terminologiques sur Internet \(TERMISTI\)](#)

[VOCOR, Vocabulaire correctif à l'intention des milieux d'éducation](#)

Ne semble pas avoir été mis à jour depuis septembre 1996.

[Glossaire de termes utilisés en production sonore, en électronique, en électroacoustique et en musique](#)

Plus de 200 termes techniques anglais et français relatifs à la production de sons, numérique comme analogique.

[CD and DVD Glossary by Media Sciences](#)

Plus de 200 termes techniques anglais relatifs au stockage sur consommables amovibles (disquettes, cédéroms, dévédéroms).

[Les logiciels gratuits pour le français et les langues anciennes recensés par l'académie de Versailles](#)

Des liens précieux, présentés avec des commentaires pertinents, bien que sous une forme peu agréable.

[Bribes](#)

Modestement sous-titré « Petit dictionnaire des citations », gratuit et sans publicité (+ de 500 pages !).

[Bases de la cuisine](#)

Pour ses trois lexiques consacrés aux termes culinaires relatifs aux bases, aux fonds et aux appareils.

[Actions pour Promouvoir le Français des Affaires](#)

Liste très complète des mots nouveaux ainsi que certains qui ne sont que conseillés.

BANQUES DE DONNÉES

PRESSE EN LIGNE

[Factiva](#)

Texte intégral des principaux titres de la presse française et internationale. Exemples : La Tribune, Les Echos, Business Week, El Pais...

[Business source premier](#)

Résumés ou texte intégral de magazines économiques dont Business Week, Forbes, Fortune

[Pressed](#)

35 millions d'articles de la presse nationale, régionale, et spécialisée françaises.

INFORMATIONS FINANCIERES SUR LES SOCIETES

[Diane](#)

Données financières sur des centaines de milliers d'entreprises françaises.

[Thomson One Banker](#)

Coordonnées et analyses financières des plus grandes sociétés cotées internationales

[Dafsaliens](#)

Filiales des 500 plus importants groupes français à travers le monde.

[Kompass](#)

160 000 entreprises françaises classées à l'aide de 50 000 codes produits

[Inpi/Jouve](#)

Brevets, marques, modèles déposés en France ou dans un pays membre de l'UE

[Questel Orbit](#)

Brevets et marques déposés au Japon et aux Etats-Unis

[Iliad](#)

Comptes sociaux, bilans, noms des dirigeants...

[Fininfo](#)

Données sur 7 millions d'entreprises y compris commerçants et artisans

MARCHES FINANCIERS

DataStream

Accès: une station près du bureau du Conseil à la Recherche.

Séries temporelles : actions, obligations, matières premières, change, indices, données macroéconomiques, profils d'entreprise.

ETUDES DE MARCHES

[Global Market Information Database \(GMID\)/ Euromonitor](#)

Données statistiques, études sectorielles et analyses macroéconomiques sur les marchés français et étranger

[XERFI](#)

Access : Sur le campus de Cergy uniquement

Etudes sectorielles couvrant les marchés français et européen.

[DELPHES](#)

Delphes permet de repérer des articles sur les produits et les marchés, parus dans des revues spécialisées.

REVUES ACADEMIQUES EN LIGNE

[Business source premier \(BSP\) / EBSCO](#)

Résumés et texte intégral de plusieurs milliers de revues dans les domaines du management et des affaires.

[ScienceDirect](#)

Texte intégral de près de 2 000 revues académiques de l'éditeur Elsevier dans tous les domaines.

[Electronic collection online \(ECO\)](#)

Texte intégral d'une quarantaine de revues de référence en anglais dans les domaines de l'économie/gestion

[EconLit](#)

Reférences d'articles académiques, d'ouvrages, de thèses, de documents de recherche de la littérature économique mondiale avec les résumés

[Emerald](#)

100 revues en texte intégral de l'éditeur Emerald dans les domaines du management, des ressources humaines, du marketing ...

[Journal Storage \(JSTOR\)](#)

Archives électroniques de revues académiques anglo-saxonnes couvrant toutes les disciplines, depuis leur premier numéro, mais pas les 5 dernières années.

RECHERCHE

[British Library Proceedings](#)

Index des publications qui font suite aux conférences, aux colloques, aux expositions, aux ateliers, etc. et qui ont été déposées à la British Library.

[CEPR Discussion Papers](#) : résumé et texte intégral des documents de recherche du Center for Economic Policy Research

[Journal Citation Reports \(JCR\)](#)

JCR permet d'évaluer et de comparer les revues académiques selon différents critères : mesure d'impact, nombre d'articles cités...

[NBER Working Papers](#)

(National Bureau of Economic Research). Texte intégral de la majorité des working papers du National Bureau of Economic Research, téléchargeables au format PDF.

[PapersFirst](#)

Reférences de tous les écrits déposés pour les congrès, les conférences, les ateliers, les symposiums et répertoriés par la British Library.

[Social Science Citation Index \(SSCI\)](#)

Index multidisciplinaire de plusieurs milliers de revues académiques en sciences sociales (depuis 1984).

REFERENCES D'ARTICLES ET D'OUVRAGES

[DELPHES](#)

Références et résumés d'articles des cinq dernières années parus dans des revues économiques et professionnelles.

[Electronic collection online \(ECO\)](#)

Références d'articles académiques en anglais dans toutes les disciplines

[EconLit](#)

Références et résumés d'articles académiques, d'ouvrages, de thèses, de documents de recherche de la littérature économique mondiale

[SUDOC \(Système universitaire de documentation\)](#)

Catalogue commun des bibliothèques de l'enseignement supérieur français.

[WorldCat](#)

Références de livres, ressources Web et autres documents du monde entier.

[FirstSearch](#)

Service donnant accès à différentes banques de données multidisciplinaires (essentiellement des références bibliographiques).

ASIE

[Reference Proquest Asia](#)

Informations sur la géopolitique, l'économie, le business de la région asiatique.

DROIT

[Jurisclasseur](#)

Fonds juridique comprenant toutes les encyclopédies JurisClasseur, la jurisprudence et la doctrine française, la législation consolidée, des revues juridiques en ligne dont la Semaine Juridique

[Lexbase](#)

Guichet unique du droit, Lexbase propose des revues juridiques, 12 bases juridiques, des sources officielles.

[Editions Francis Lefebvre](#)

Mémentos, jurisprudence, doctrine...

[Lamy/Wolters Kluwer](#)

Ouvrages et revues de l'éditeur avec 85 millions d'hyperliens

ACCES MULTIBASES

[Thomson](#)

250 bases. Énergie, défense, marketing, biotechnologie

[ReedElsevier](#)

36.000 sources d'information. Droit, finance, économie, presse.

MOTEURS DE RECHERCHE INCONTOURNABLES / ORIGINAUX

Google

MSN

Yahoo !

AskJeeves

Exalead

Grokker (visualisation graphique des résultats de Yahoo !)

Marumushi (résultats sous forme de mosaïque)

ANNUAIRES PROFESSIONNELS

Abc-Luxe.com

Bepub.com

Anvar.fr

Veille.com

Abc-Annuaire.com

Qualisteam.com

Virtualpet.com/industry

Iae-Paris.org

Rime.ccip.fr

Apce.com

Furl.net

Amf-france.org

Acfc.cci.fr

Centre-affaires.net

Lesannuaires.com

Indicateur.com

Rime

Qualisteam.com

BLOGS A USAGE PROFESSIONNEL

www.marketorama.typepad.com

www.nouveaunjour.fr

www.sepulveda.net

<http://pascal.blogs.com>

www.gizmodo.com

www.lebloggadget.com

<http://oseres.typepad.com>

www.moovement.com

www.altaide.fr

<http://prplanet.typepad.com/ceobloggersfrench>

<http://bcouly.blogs.com>

FORUMS DE DISCUSSION

<http://forum.netpme.fr>
<http://boursorama.com>
<http://forum.actufinance.fr>
www.laviefinanciere.com
<http://forum.presence-pc.com>
<http://forum.hardware.fr/>
<http://grenouille.com>
<http://homecinema-fr.com>
<http://forum-auto.com>
www.aufeminin.com/
www.ciao.fr

LECTEURS RSS

[Alerteinfo](#)
[ActivRefresh](#)
[NewsGator](#)
[Guru News Ticker](#)