



REFERENCEMENT : STRATEGIE DOCUMENTAIRE VERSUS STRATEGIE MARKETING Le cas des sites web des cyberlibrairies et maisons d'édition

Adeline Marck

► To cite this version:

Adeline Marck. REFERENCEMENT : STRATEGIE DOCUMENTAIRE VERSUS STRATEGIE MARKETING Le cas des sites web des cyberlibrairies et maisons d'édition. domain_shs.info.docu. 2005. mem_00526900v1

HAL Id: mem_00526900

https://memsic.ccsd.cnrs.fr/mem_00526900v1

Submitted on 16 Oct 2010 (v1), last revised 25 Jan 2011 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CONSERVATOIRE NATIONAL DES ARTS ET METIERS
INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION

MEMOIRE
pour obtenir le
DESS en Sciences de l'information et de la documentation spécialisées

présenté et soutenu par
Adeline MARCK

le 25 octobre 2005

**REFERENCEMENT : STRATEGIE DOCUMENTAIRE
VERSUS STRATEGIE MARKETING**
Le cas des sites web des cyberlibrairies et maisons d'édition

Membres du jury :
Madame Danièle DEGEZ
Madame Patricia MACAIGNE
Monsieur José GARCIA

Cycle supérieur Promotion XXXV

Remerciements

Je tiens à exprimer toute ma gratitude à Madame Danièle DEGEZ - qui a accepté de suivre l'évolution de ce travail - et à Monsieur José Garcia, auprès duquel j'ai appris avec plaisir le métier de documentaliste « de terrain » lors d'un stage formateur à l'ICOMOS.

Ce mémoire clôt une année riche en rencontres et en échanges fructueux. Bonne chance à tous les élèves du groupe 1 ...

A toi Patrice, que la gentillesse et l'humour ne quittent jamais, merci pour tes conseils avisés d'ex-référenceur !

Christine et Virginie, je suis heureuse d'avoir parcouru un petit bout de chemin avec vous. Alors, ce bon restaurant, c'est pour quand ? La fin d'un tel marathon mérite bien une petite coupe de Champagne !

Je remercie enfin mes parents pour leur infinie patience - c'est promis, cette fois, j'arrête ! ;-)
– et mon ami, pour ne pas m'avoir noyé trop vite sous les travaux agricoles et de BTP...

Résumé

Notice : MARCK, Adeline. *Référencement : stratégie documentaire versus stratégie marketing. Le cas des sites web des cyberlibrairies et maisons d'édition*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 2005, 170 p.

Résumé : L'impact d'un site (ou d'une page) Web dépend étroitement de sa visibilité sur Internet, autant dire de son positionnement sur les principaux moteurs de recherche. C'est pourquoi le référencement se situe au cœur de la stratégie de communication des sites de e-commerce.

Après une présentation générale des moteurs de recherche et de leur fonctionnement, sont abordées les méthodes d'optimisation des pages web en vue de leur référencement. Puis sont définis les concepts fondamentaux liés à cette opération, et clarifié son double objectif documentaire et marketing. Le troisième chapitre replace le référencement dans le cadre plus vaste de la promotion d'un site web et détaille les méthodes permettant d'évaluer sa qualité. Une analyse comparative des codes sources, et en particulier des métadonnées de sept sites Internet de cyberlibrairies et maisons d'édition, complète cette étude.

Des bibliographies (analytique, alphabétique, complémentaire), une webographie, un glossaire et un index viennent clore ce mémoire.

Descripteurs : COMMUNICATION PUBLICITAIRE, ETUDE COMPARATIVE, INTERNET, MARKETING, METADONNEE, MOTEUR DE RECHERCHE, OPTIMISATION, OUTIL DE RECHERCHE, PAGE WEB, REFERENCEMENT, REFERENCEMENT PAYANT, SITE WEB

Table des matières

<i>Remerciements</i>	<i>1</i>
<i>Résumé</i>	<i>2</i>
<i>Table des matières.....</i>	<i>3</i>
<i>Liste des tableaux.....</i>	<i>9</i>
<i>Liste des annexes.....</i>	<i>10</i>
 INTRODUCTION	 12
 CHAPITRE 1 - Les moteurs de recherche	 15
1. Qu'est-ce qu'un moteur de recherche ?	16
1.1. Présentation des outils de recherche.....	16
1.1.1. Différence entre annuaire et moteur de recherche	16
○ Annuaire	16
○ Moteur de recherche	16
○ Méta-moteur	17
1.1.2. Les principaux outils de recherche francophones	17
1.2. Principe de fonctionnement d'un moteur de recherche.....	18
1.3. Systèmes d'indexation.....	19
1.3.1. Indexation par les moteurs de recherche	19
1.3.2. Différence entre indexation documentaire et indexation informatique.....	20
1.4. Méthodes de tri des résultats et <i>ranking</i>	22
1.4.1. Critères de pertinence liés au contenu de la page (« <i>in -page</i> »).....	22
○ Contenu de la balise <TITLE>	22
○ Balises <META>.....	22
1) Les meta NAME	23
2) Les meta HTTP-EQUIV	23
3) Les meta Dublin Core	23
○ Texte visible de la page	24
○ Attribut ALT de la balise 	25
○ Contenu de l'URL	25
○ Autres critères.....	25
○ Indice de densité d'un mot-clé	25
1.4.2. Critères de pertinence liés au site (« <i>off-page</i> »).....	26
○ Nom de domaine.....	26
○ Indice de popularité d'une page (IPP).....	26
○ Indice de clic.....	26

1.4.3. Classement des pages (<i>ranking</i>)	27
○ Tri par pertinence	27
○ Tri par popularité	27
○ Tri par mesure d'audience	28
○ Tri par catégories	28
1.5. Limites des moteurs de recherche	29
2. La recherche avec un moteur	30
2.1. Moteur d'interrogation	30
2.2. Type de recherche	30
2.3. Page de résultats	31
2.4. Sémantique et syntaxe d'interrogation	31
2.5. Questions générales ou ciblées	31
CHAPITRE 2 - Le référencement : une technique documentaire ou marketing ?	33
1. Intérêts et enjeux du référencement	34
1.1. Définition	34
1.2. De l'importance d'être référencé sur les moteurs de recherche	34
1.3. Du référencement au positionnement	34
1.4. Différence entre indexation et référencement	35
2. Optimiser une page Web pour le référencement	36
2.1. Les règles générales de l'optimisation	36
2.2. Sélection des mots-clés	36
2.2.1. Définition	37
2.2.2. Choix des mots-clés	37
2.3. Facteurs bloquants	39
2.3.1. Les <i>frames</i>	39
2.3.2. Flash et Javascript	39
2.3.3. Les pages dynamiques	39
3. Un référencement différencié en fonction des outils de recherche et des objectifs	40
3.1. Référencement auprès des annuaires	40
3.2. Référencement classique auprès des moteurs de recherche	41
3.2.1. Référencement naturel	41
3.2.2. Référencement par indexation	41
3.2.3. Référencement par positionnement	41

3.3. Quand le référencement est devenu de la publicité.....	42
3.3.1. Liens sponsorisés.....	43
3.3.2. Liens contextuels.....	44
3.3.3. <i>Pay Inclusion</i>	45
○ L'inclusion payante d'URL.....	45
○ Le flux XML ou <i>Trust Feed</i>	45
3.4. Le référencement des pages dynamiques.....	46
3.5. Référencement manuel ou référencement automatique ?	47
4. Dérives du référencement.....	48
4.1. La vague du payant.....	48
4.2. Le <i>spamdexing</i>	48
4.3. Le positionnement : une perversion de l'indexation ?	49
5. Les sociétés spécialisées dans le référencement.....	50
 CHAPITRE 3 - Le référencement : une étape dans le plan marketing d'un site web	51
1. Le site web : un produit.....	52
2. Les préalables au référencement.....	53
2.1. Analyse de l'environnement.....	53
2.1.1. Analyse des sites concurrents.....	53
2.1.2. Connaître les habitudes des internautes	54
2.2. Choix de mots-clés pertinents en fonction de la cible.....	54
3. Analyses et mesures d'audience.....	56
3.1. Mesures côté client (<i>user-centric</i>).....	56
3.2. Mesures côté serveur (<i>site-centric</i>)	56
3.2.1. Analyse des logs.....	56
3.2.2. Analyse à distance.....	57
3.2.3. Analyse des <i>sniffers</i>	57
3.3. Estimation du trafic	57
3.3.1. Nombre de hits	57
3.3.2. Nombre de pages vues.....	58
3.3.3. Nombre de visites ou nombre de sessions.....	58
3.3.4. Nombre de visiteurs	58
3.4. Identification du profil des internautes.....	58
3.5. Analyse du comportement des internautes.....	59
3.6. Performance et notoriété d'un site web.....	59
3.6.1. Accès direct.....	60
3.6.2. Accès par liens externes	60
3.6.3. Accès par les outils de recherche	60

4. Suivi du référencement et de l'indexation.....	61
4.1. Détermination <i>a posteriori</i> des mots-clés	61
4.2. Suivi du positionnement.....	61
4.3. Rapports de positionnement	62
5. Le référencement : une méthode parmi d'autres dans le processus de promotion d'un site web.....	63
5.1. Campagnes <i>on-line</i>	63
5.1.1. Affiliation	63
5.1.2. Cybermarketing.....	63
5.1.3. Fils RSS.....	64
5.1.4. Publicité en ligne	64
5.1.5. <i>Webring</i> ou chaîne de sites.....	64
5.1.6. Autres moyens de promotion en ligne.....	65
5.2. Campagnes <i>off-line</i>	65
CHAPITRE 4 - Etude comparative : sites web de cyberlibrairies et maisons d'édition.....	66
1. Recueil des informations.....	67
2. Présentation des sociétés et sites étudiés	68
2.1. Les cyberlibrairies	68
2.1.1. Alapage.fr.....	68
2.1.2. Amazon.fr.....	68
2.1.3. Chapitre.com	69
2.1.4. Fnac.com	69
2.2. Les maisons d'édition.....	69
2.2.1. Dunod.com	69
2.2.2. Eyrolles.com.....	70
2.2.3. Lavoisier.com.....	70
2.3. Des sites dynamiques	70
3. Optimisation des pages.....	72
3.1. L'omniprésence d'Amazon : un positionnement stratégique.....	72
3.1.1. Méthodologie	72
3.1.2. Résultats	73
3.1.3. Comparatif des sites les mieux référencés	73
3.2. L'optimisation des pages.....	74
3.2.1. Contenu de la balise <TITLE>	75
3.2.2. Le recours aux métadonnées	76
○ Meta Description	77
○ Meta Keywords	78
○ Meta NAME	79
○ Meta HTTP-EQUIV	80

3.2.3. Texte visible de la page	81
3.2.4. Contenu invisible de la page	81
○ Facteurs bloquants	81
○ Balise ALT	82
3.2.5. Mise en exergue et valeur ajoutée	82
3.2.6. Contenu de l'URL	83
3.2.7. Indice de densité des mots-clés	84
4. Visibilité et positionnement des sites sur les outils de recherche	85
4.1. Visibilité des pages.....	85
4.2. Nombre de pages indexées	86
4.3. PageRank© de Google	86
4.4. Indice de popularité	87
5. Le plan marketing des sites étudiés	89
5.1. Liens sponsorisés.....	89
5.2. Affiliation	90
5.3. Des sites supports de publicité	91
5.4. Fils RSS.....	91
CONCLUSION.....	94
BIBLIOGRAPHIE	96
Présentation de la bibliographie	97
Bibliographie analytique	99
Bibliographie alphabétique	113
Bibliographie complémentaire	121
WEBOGRAPHIE	126
GLOSSAIRE.....	132

ANNEXES	139
<i>Annexe 1 : capture d'écran d'une page produit Alapage.com</i>	<i>140</i>
<i>Annexe 2 : en-tête HTML d'une page produit Alapage.com.....</i>	<i>141</i>
<i>Annexe 3 : capture d'écran d'une page produit Amazon.fr</i>	<i>143</i>
<i>Annexe 4 : en-tête HTML d'une page produit Amazon.fr</i>	<i>146</i>
<i>Annexe 5 : capture d'écran d'une page produit Chapitre.com.....</i>	<i>148</i>
<i>Annexe 6 : code source HTML d'une page produit Chapitre.com</i>	<i>149</i>
<i>Annexe 7 : capture d'écran d'une page produit Fnac.com.....</i>	<i>151</i>
<i>Annexe 8 : en-tête HTML d'une page produit Fnac.com.....</i>	<i>152</i>
<i>Annexe 9 : capture d'écran d'une page produit Dunod.com.....</i>	<i>154</i>
<i>Annexe 10 : en-tête HTML d'une page produit Dunod.com</i>	<i>155</i>
<i>Annexe 11 : capture d'écran d'une page produit Eyrolles.com</i>	<i>156</i>
<i>Annexe 12 : en-tête HTML d'une page produit Eyrolles.com.....</i>	<i>158</i>
<i>Annexe 13 : capture d'écran d'une page produit Lavoisier.fr.....</i>	<i>159</i>
<i>Annexe 14 : code source HTML d'une page produit Lavoisier.fr</i>	<i>160</i>
<i>Annexe 15 : positionnement des sites étudiés sur les moteurs de recherche</i>	<i>161</i>
<i>Annexe 16 : texte visible des pages produits.....</i>	<i>162</i>
<i>Annexe 17 : densité des mots trouvés dans le texte visible des pages produits</i>	<i>163</i>
<i>Annexe 18 : liens sponsorisés sur certains mots-clés</i>	<i>164</i>
 INDEX ALPHABETIQUE	 165

Liste des tableaux

Tableau 1 : “Top 10” des moteurs de recherche en fonction des parts de trafic généré (Juillet 2005).....	17
Tableau 2 : Taille des index de six moteurs de recherche	20
Tableau 3 : Liste des annexes fournies pour chaque site étudié.....	67
Tableau 4 : Technologies mises en œuvre par les sites étudiés	71
Tableau 5 : Qui est le mieux référencé pour le mot-clé Saporta référencement sur le net ?... 73	
Tableau 6 : Qui est le mieux référencé pour le mot-clé check-list pour réussir son site web ?	74
Tableau 7 : Qui est le mieux référencé pour le mot-clé Marie Prat référencement de votre site web ?	74
Tableau 8 : Titre des pages produit	75
Tableau 9 : Meta description – Pages produits	77
Tableau 10 : Meta Keywords – Pages produits	78
Tableau 11 : Emploi des meta name dans les pages produits	79
Tableau 12 : Emploi des meta HTTP-EQUIV dans les pages produits	80
Tableau 13 : Contenu des pages produits.....	81
Tableau 14 : Mise en exergue du contenu.....	82
Tableau 15 : URL des pages produits.....	83
Tableau 16 : Classement de visibilité des pages produits en fonction des requêtes.....	85
Tableau 17 : Nombre de pages indexées.....	86
Tableau 18 : PageRank© Google des sites étudiés	86
Tableau 19 : Indice de popularité des sites étudiés	87

Liste des annexes

<i>Annexe 1 : capture d'écran d'une page produit Alapage.com</i>	<i>140</i>
<i>Annexe 2 : en-tête HTML d'une page produit Alapage.com.....</i>	<i>141</i>
<i>Annexe 3 : capture d'écran d'une page produit Amazon.fr</i>	<i>143</i>
<i>Annexe 4 : en-tête HTML d'une page produit Amazon.fr</i>	<i>146</i>
<i>Annexe 5 : capture d'écran d'une page produit Chapitre.com.....</i>	<i>148</i>
<i>Annexe 6 : code source HTML d'une page produit Chapitre.com</i>	<i>149</i>
<i>Annexe 7 : capture d'écran d'une page produit Fnac.com.....</i>	<i>151</i>
<i>Annexe 8 : en-tête HTML d'une page produit Fnac.com.....</i>	<i>152</i>
<i>Annexe 9 : capture d'écran d'une page produit Dunod.com.....</i>	<i>154</i>
<i>Annexe 10 : en-tête HTML d'une page produit Dunod.com</i>	<i>155</i>
<i>Annexe 11 : capture d'écran d'une page produit Eyrolles.com</i>	<i>156</i>
<i>Annexe 12 : en-tête HTML d'une page produit Eyrolles.com.....</i>	<i>158</i>
<i>Annexe 13 : capture d'écran d'une page produit Lavoisier.fr.....</i>	<i>159</i>
<i>Annexe 14 : code source HTML d'une page produit Lavoisier.fr</i>	<i>160</i>
<i>Annexe 15 : positionnement des sites étudiés sur les moteurs de recherche</i>	<i>161</i>
<i>Annexe 16 : texte visible des pages produits.....</i>	<i>162</i>
<i>Annexe 17 : densité des mots trouvés dans le texte visible des pages produits</i>	<i>163</i>
<i>Annexe 18 : liens sponsorisés sur certains mots-clés</i>	<i>164</i>

Conventions typographiques

Les références complètes sont présentées entre crochets sous la forme normalisée suivante :

[Auteur, année, p.xx, (n° de renvoi)]

Elles comprennent : le nom de l'auteur (personne physique ou morale), l'année de publication de l'article ou de l'ouvrage et la pagination, les chiffres entre parenthèses correspondant au numéro de la référence dans la bibliographie analytique (p. 99-112). Ces références renvoient d'autre part à la bibliographie alphabétique des noms d'auteurs (p. 113-120).

Les citations sont reconnaissables à leur texte en italique placé entre guillemets.

Les termes et expressions qui font l'objet d'une définition dans le glossaire (p. 132-138) sont suivis d'une astérisque (*) lors de leur première apparition dans le texte.

INTRODUCTION

Le Web draine un flux croissant d'information véhiculé par un nombre de sites en constante augmentation. Comment, dans ces conditions, faire connaître l'existence d'un site et lui assurer une visibilité suffisante pour qu'il soit visité par les internautes ?

Un site n'a d'existence que s'il est connu. Son adresse doit donc être largement diffusée. Sa promotion passe entre autre par le référencement*, c'est-à-dire l'inscription du site sur les outils de recherche* (moteurs*, annuaires* et portails spécialisés).

Le référencement, étroitement lié à la notion d'indexation*, une pratique documentaire reconnue, peut se concevoir à la fois comme une démarche typiquement documentaire, visant à faciliter l'accès à un site web, et comme un outil marketing chargé d'accroître la visibilité et le positionnement* d'un site sur les outils de recherche pour en augmenter le trafic* et donc la rentabilité.

C'est pour cerner au plus près la nature du référencement et ses implications que nous avons lancé cette étude.

Le traitement de cette question exige un examen attentif du fonctionnement des moteurs de recherche de même qu'une connaissance des techniques, comme nous le verrons variées, de référencement, sans lesquelles ne peut se construire une stratégie marketing plus globale sur Internet.

A l'issu d'un examen de la littérature, une étude permettra de comparer les stratégies retenues par des acteurs du monde de l'information. Sur la Toile*, la visibilité des maisons d'édition est atténuée par l'omniprésence de grosses structures de e-commerce concurrentes comme Amazon, Alapage ou la Fnac, lesquelles raflent la plupart des commandes en ligne. Comment font-elles pour se retrouver dans les premières positions lors des requêtes* ciblées sur les livres ? Quelles sont leurs stratégies de référencement et de positionnement : se fondent-elles sur une logique documentaire ou sur une démarche marketing agressive ?

Ballottés entre deux objectifs, celui de faciliter l'accès à l'information tout en honorant des impératifs commerciaux, les sites web des cyberlibrairies et maisons d'édition apparaissent comme des sujets d'étude particulièrement adaptés à la problématique qui nous occupe.

Nous avons concentré notre analyse sur deux thèmes très précis :

1. le référencement sur les moteurs de recherche (visibilité, positionnement, optimisation*)
2. la stratégie marketing adoptée (liens sponsorisés* et contextuels*, techniques de promotion).

La méthode retenue pour appréhender la nature du référencement de ces différentes entités s'appuie sur l'utilisation d'utilitaires disponibles sur le web et l'analyse des codes sources HTML* des pages, le questionnaire envoyé pour l'occasion et censé éclairer les techniques de référencement et la stratégie marketing des sites étudiés étant resté lettre morte.

CHAPITRE 1

Les moteurs de recherche

1. Qu'est-ce qu'un moteur de recherche ?

Trois possibilités s'offrent à l'internaute pour trouver un site : soit il connaît le nom de domaine* et le tape directement dans son navigateur*, soit le lien* est enregistré dans ses favoris. Il peut aussi utiliser des sites où se retrouve concentrée et organisée l'information : les outils de recherche [Carlier, 2000, p. 170, **(01)**]. Ceux-ci peuvent être généralistes ou spécialisés sur un thème donné.

1.1. Présentation des outils de recherche

1.1.1. Différence entre annuaire et moteur de recherche

○ Annuaire

Appelé également guide ou répertoire, un annuaire est « *un catalogue de sites, hiérarchisés par catégories ou rubriques.* » [Carlier, 2000, p. 171, **(01)**]. Les thèmes sont organisés du plus général au plus particulier. Il est constitué d'un ensemble de fiches descriptives présentant l'URL*, le titre et un résumé du site indexé, parfois son origine géographique, le public visé et la langue employée. Des « documentalistes », également appelés « netsurfeurs », sont chargés de visiter, évaluer puis classer les sites soumis dans les rubriques les plus appropriées. Ils se réservent le droit de modifier les données transmises (commentaires, titre, rubriques proposées) [Andrieu, 2000, p. 348, **(02)**].

Les annuaires proposent deux modes d'interrogation : (1) via le mode de navigation hiérarchique dans la classification (catégories, sous-catégories, rubrique, liste) ; (2) sous forme de requête à l'image des moteurs de recherche.

Parmi les principaux annuaires francophones, nous pouvons citer le Guide de Voila, l'Open Directory Project, Lycos France, Yahoo! France et Nomade.

○ Moteur de recherche

Un moteur de recherche est un outil constitué d'un ensemble de robots* logiciels (appelés agents*, aspirateurs, araignées* ou robots en français et *spiders**, *wanderers*, *crawlers** en anglais).

Ces programmes de navigation parcourent le Web de lien en lien de manière continue en indexant automatiquement le contenu des pages visitées [Chu, 2003, p. 96, **(03)** ; Andrieu, 2005, p. 8-10, **(04)**].

○ Méta-moteur

Les outils de recherche d'information, méta-moteurs* (Metacrawler, Kartoo)¹ et logiciels (Copernic, Strategic Finder), interrogent simultanément une sélection de moteurs. Ils classent et exposent les résultats de ces recherches sur leur propre interface.

1.1.2. Les principaux outils de recherche francophones

D'après une étude de l'institut Mediamétrie², la France comptait 24,3 millions d'internautes soit 46,9 % de la population âgée de 11 ans et plus en juillet 2005. Cette même étude révèle la présence de quelques moteurs de recherche parmi le « Top 30 français » des sites les plus consultés (hors applications)³ dont, en tête de liste Google, suivi de MSN (3^e) et Yahoo ! (6^e). La société de référencement 1^{ère} position propose chaque mois ses analyses sur les statistiques fournies par l'outil de mesure d'audience Xiti. Celles-ci permettent de dresser le « Top 10 » des moteurs de recherche majeurs sur le marché, dont une vision synthétique est proposée à travers ce tableau :

Tableau 1 : “Top 10” des moteurs de recherche en fonction des parts de trafic généré (Juillet 2005)

Moteur	Parts de trafic généré – France – Xiti-1ère position – Edition de juillet 2005	URL
1. Google	78,47 %	http://www.google.fr
2. Yahoo !	6,31 %	http://www.yahoo.fr
3. Voilà	5,41 %	http://www.voila.fr
4. MSN France	3,77 %	http://www.msn.fr/
5. AOL.fr	2,04 %	http://www.aol.fr
6. Free	1,36 %	http://www.free.fr
7. AltaVista	0,41 %	http://www.altavista.fr
8. Nomade	0,40 %	http://www.nomade.tiscali.fr
9. Club Internet	0,36 %	http://www.club-internet.fr
10. Lycos	0,26 %	http://www.lycos.fr

Source : <http://www.secrets2moteurs.com/barometre/> - Consulté le 22 août 2005

¹ Voir <http://www.indicateur.com/metamoteur> [consulté le 31 août 2005]

² L'audience de l'Internet en France en Juillet 2005. -

http://www.mediametrie.fr/resultats.php?resultat_id=177&rubrique=net [consulté le 26 août 2005]

³ Les applications se repèrent à leur extension en .exe

Google et Yahoo ! représentent à eux seuls 84,78 % du marché français, le premier demeurant toutefois le leader incontesté des moteurs de recherche sur le Web avec 78,47 % du volume total généré en juillet 2005. Les autres moteurs de recherche grappillent quelques internautes mais leur part reste dérisoire face à la suprématie de leur puissant concurrent.

Les moteurs de recherche ont relégué les annuaires (Yahoo!⁴, Voilà⁵, Lycos⁶) aux oubliettes, ces derniers ne générant qu'un trafic ridicule. Seul l'Open Directory Project⁷, utilisé par Google résiste encore. Il faut dire qu'il est basé sur un modèle singulier : les évaluateurs sont tous bénévoles.

1.2. Principe de fonctionnement d'un moteur de recherche

Le robot d'un moteur de recherche est en fait « [...] *un programme qui effectue des recherches automatisées sur Internet. Il parcourt l'ensemble des fichiers constituant un site, emmagasine les informations dans l'index* et classe les mots-clés* et le texte de chaque page qu'il trouve.* » [Chu, 2003, p. 96, (03)].

Le fonctionnement d'un moteur de recherche se découpe donc en trois phases, que nous allons, dès à présent, détailler :

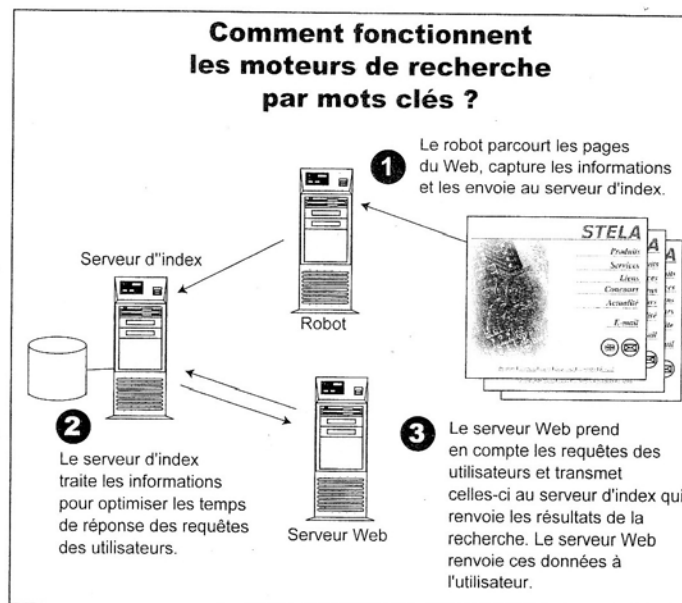
- 1) les robots explorent le Web et collectent des informations.
- 2) les robots transmettent les données à l'indexeur. Ce dernier est chargé d'extraire les informations nécessaires à l'indexation tels que les éléments visibles de la page, les mots-clés, les liens et autres métadonnées* [Calishain & Dornsfet, 2003, p. 318, (05)].
- 3) le site est inséré dans la base de données du moteur. Suite à une requête effectuée sur une interface de recherche, les résultats sont restitués sous la forme d'une liste ordonnée où ils sont classés en fonction de leur pertinence (*ranking*) [Andrieu, 2005, p. 9, (04)].

⁴ <http://fr.search.yahoo.com/dir> [consulté le 22 août 2005]

⁵ <http://guide.voila.fr> [consulté le 22 août 2005]

⁶ <http://www.lycos.fr> [consulté le 22 août 2005]

⁷ <http://www.dmoz.org> [consulté le 22 août 2005]



Comment fonctionnent les moteurs de recherche par mots-clés ?

La capture des pages se déroule en trois phases : le robot parcourt le web et envoie les pages au serveur d'index qui sert de référence au serveur web abritant l'interface utilisateur.

Source : Andrieu, 2000, p. 313, (02), avec son aimable autorisation

Les robots ne traitent pas tous les sites de la même manière. Ils explorent plus fréquemment les sites à fort trafic ou à fort taux de renouvellement des contenus aux dépens des pages « statiques » [Andrieu, 2005, p. 10, (04)].

1.3. Systèmes d'indexation

L'indexation des pages peut s'effectuer de deux manières : par indexation automatique* en texte intégral* ou par déclaration volontaire auprès du moteur de recherche par le biais d'un formulaire [Andrieu, 2000, p. 314, (02)].

1.3.1. Indexation par les moteurs de recherche

La procédure d'indexation s'effectue en deux temps : le robot n'indexe d'abord que les pages soumises (*fresh crawler* de Google) puis il effectue l'indexation complète du site (*deep crawler* ou *full crawler* de Google) (ce qui peut prendre plusieurs semaines ou plusieurs mois) [Andrieu, 2000, p. 318, (02)].

Les informations récoltées par le robot sont ensuite transmises au moteur d'indexation. Les pages collectées ne sont cependant pas toutes retenues.

Des critères comme la richesse du contenu et sa lisibilité par le robot, l'adéquation entre les mots-clés présents dans les balises meta et le contenu des pages influent sur l'indexation définitive. A l'issue de cette étape, la taille de l'index est majorée des nouvelles données glanées et traitées par le robot. Le tableau ci-dessous permet de comparer la taille des index de plusieurs moteurs de recherche :

Tableau 2 : Taille des index de six moteurs de recherche

	Exalead	Gigablast	Ask Jeeves	MSN	Google	Yahoo!
Nb de pages indexées	+ de 1 milliard	1,5 milliards	2,5 milliards	5 milliards	+ de 8 milliards	20 milliards
Source	Exalead	Gigablast	Search EngineWatch	MSN	Google	Yahoo
Date	02/2005	03/2005	11/2004	02/2005	11/2004	08/2005

Source : Andrieu, 2005, p. 12, (04) à l'exception des données sur Yahoo : Andrieu, 2005, (06).

L'index est constitué d'un index principal (ou "maître"), qui contient tout le corpus d'informations puisé par le robot, ainsi que des fichiers inverses* ou index inversés, regroupant tous les mots-clés attachés aux adresses (URL) des documents comprenant ces termes [Andrieu, 2005, p. 11, (04)].

1.3.2. Différence entre indexation documentaire et indexation informatique

L'indexation est une démarche fondamentale dans la pratique documentaire : « *elle représente l'opération centrale de traitement des informations et des documents [...].* » [Cacaly, 2004, p. 114-115, (07)]. La norme* AFNOR (Association Française de Normalisation) Z 47-102 (1993) définit l'indexation documentaire comme « *l'opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire* les concepts après les avoir extraits du document par une analyse. [...]* » [AFNOR, 2000, p. 396, (08)].

Mais cette notion diffère en fonction du domaine concerné. En effet, l'indexation documentaire, manuelle et basée sur un vocabulaire contrôlé, et l'indexation informatique, automatique en texte intégral, sont loin de partager les mêmes modes de traitement des documents et ne qualifient pas la même procédure.

D'un point de vue documentaire, l'indexation « *est une opération intellectuelle impliquant une analyse approfondie d'un document et la représentation condensée de l'information portée par ce document.* » [Chaumier, 2000, p. 14, **(09)**]. Il s'agit de traduire les concepts fondamentaux d'un document en un ensemble de termes caractéristiques ou « mots-clés ». Le processus se déroule en deux phases : une phase d'analyse conceptuelle et une phase de traduction ou de représentation des concepts dans un langage documentaire.

D'un point de vue informatique, l'indexation définit l'« *extraction par un ordinateur des unités lexicales qui permettent de déterminer et de caractériser le contenu d'un document textuel* »⁸. Ce processus automatisé indexe en texte intégral (ou *full text* en anglais). Ses avantages - disponibilité rapide, accès multiple à l'information (pas de restriction de mots-clés), vocabulaire riche (pas de normalisation réductrice), absence de subjectivité [Chirié, 1995, p. 24-25, **(10)**] - lui confèrent une souplesse indéniable. Toutefois, l'indexation en texte intégral n'est pas non plus exempte d'inconvénients : l'absence de représentation conceptuelle des documents et les difficultés liées à la langue (synonymie, langage imagé, analogies, polysémie, homographie, etc.) s'ajoutent à la piètre qualité de l'indexation dans la mesure où tous les mots sont clés de recherche potentielles [Andrieu, 2001, **(11)**]. L'utilisation des procédés statistiques (calcul de cooccurrences, fréquence d'apparition des mots, pondération des termes, affectation d'indices, etc.) et les traitements linguistiques (analyse morphologique et syntaxique) ont cependant permis d'améliorer les performances de l'indexation en texte intégral [Urie, 2004, p. 14, **(12)**].

La notion d'indexation en informatique n'a rien à voir avec le concept d'indexation au sens traditionnel du terme. C'est une opération mécanique, à la différence de l'indexation documentaire, qui est avant tout un processus intellectuel.

⁸ Indexation automatique. <http://www.ebsi.umontreal.ca/termino/00000134.htm> [consulté le 7 septembre 2005]

1.4. Méthodes de tri des résultats et *ranking*

Plusieurs critères, propres à chaque moteur, sont utilisés pour déterminer le classement d'une page Web.

1.4.1. Critères de pertinence liés au contenu de la page (« *in-page* »)

Les critères de pertinence « *in-page* » - c'est-à-dire liés au contenu de la page - vont maintenant être évoqués plus en détail.

○ Contenu de la balise <TITLE>*

Elle contient le libellé du titre. Pour la majorité des moteurs, c'est le critère de pertinence principal [Andrieu, 2001, p. 98, (11)]. C'est la raison pour laquelle cette balise* nécessite un maximum de soin et d'attention. Les termes les plus importants doivent être présents et placés au début du titre (idéalement dans les dix premiers mots). En effet, Sullivan [Sullivan, 2002, (13)] a remarqué que la médiocrité du positionnement d'un site, pourtant pertinent au regard de la requête, résulte souvent de l'absence de mots-clés dans le titre. Toutefois, ce dernier ne peut se résumer à une succession de termes : sa formulation doit être explicite et révélatrice du contenu du document car il a pour fonction « d'appâter » les internautes. C'est sur son descriptif qu'ils jugeront de la pertinence du site.

Vu son importance, chaque page du site requiert un titre propre et conforme à son contenu [Andrieu, 2000, p. 187, (02)].

○ Balises <META>

C'est en 1969 que le terme « metadata » fut employé pour la première fois, dans un sens désormais fort éloigné du concept actuel. Il faut attendre les années 1990 pour voir apparaître les vrais standards* de métadonnées sous l'impulsion des agences gouvernementales américaines chargées de l'information géospatiale [Amerouali, 2001, p. 44, (14)].

« Les métadonnées (ou metadata) peuvent être définies comme étant des données relatives à d'autres données (data about data : données sur des données). » [Amerouali, 2001, p. 43, (14)]. On distingue trois types de métadonnées : descriptives, structurales et administratives [Morel-Pair, 2004, (15)].

Sur le web, les balises <META> ont pour fonction de faciliter l'identification et l'extraction de l'information par les moteurs. On rencontre principalement trois grandes familles de *meta tags** placées dans l'en-tête (<head>*) des pages HTML : les meta name, les meta HTTP-EQUIV et les meta Dublin Core*.

1) Les meta NAME

Ce sont les métadonnées les plus utilisées, elles permettent de décrire une page HTML.

La balise meta « Description »* : elle contient une phrase résumant le contenu de la page. Sa longueur idéale se situe entre 150 et 200 caractères, espaces compris. Cette balise est parfois indexée par les moteurs et prise en compte dans le classement de pertinence [Andrieu, 2000, p. 191-192, **(02)**]. Le résumé qui accompagne le titre des pages affichées dans les résultats lors d'une requête est souvent tiré du texte contenu dans cette balise. En revanche, Google l'ignore et génère automatiquement sa propre description (récupérée sur l'annuaire Dmoz pour les sites qui y sont inscrits).

La balise meta « Keywords »* : cette balise fournit une liste de mots-clés ciblés pouvant indiquer certains mots importants non contenus dans le document. Ils sont séparés par une virgule, un espace ou une virgule suivie d'un espace. Le nombre idéal de mots-clés se situe entre vingt et trente, en tous cas moins de mille caractères [Craven, 2004, p. 269, **(16)**]. Les moteurs de recherche principaux dont Google n'en tiennent plus compte dans le classement des pages. Cependant, elle présente un intérêt puisqu'elle permet de saisir toutes les variantes orthographiques (pluriel/singulier, masculin/féminin, etc.), dont les erreurs, d'un mot-clé [Andrieu, 2005, **(17)**].

2) Les meta HTTP-EQUIV

Peu courantes, peu utilisées, elles permettent d'envoyer des informations supplémentaires au navigateur via le protocole HTTP*.

3) Les meta Dublin Core

Roland Dachelet décrit le Dublin Core comme « [...] *le plus petit dénominateur commun des formats de description* [...] », « [...] *une sorte de lingua franca de l'indexation.* » [Dachelet, 1998, p. 185, **(18)**].

Le standard de métadonnées Dublin Core permet de décrire, de façon souple, les ressources disponibles sur le Web. Il a été créé en 1994 lors d'un colloque international du W3C* (*World Wide Web Consortium*) organisé à Dublin dans l'état de l'Ohio (USA) auquel participaient des professionnels des bibliothèques et de l'informatique [Amerouali, 2001, p. 44, **(14)**]. Le Dublin Core non qualifié, dans sa version la plus simple, comprend quinze éléments, définis chacun selon dix attributs (norme ISO/IEC 11179). Le Dublin Core qualifié se compose non seulement des éléments de base mais aussi des « qualificatifs » qui viennent affiner leur signification [Morel-Pair, 2004, **(15)**]. On peut regrouper ces éléments de base en trois catégories : la description du contenu de la ressource (Titre, Sujet/Mots-clés, Description, Source, Langue, Relation, Couverture), l'indication de la propriété intellectuelle (Auteur/Créateur, Editeur, Collaborateur, Droits) et l'identification de la ressource (Date, Type, Format, Identificateur) [Bureau & Ebacher, 2000, p. 25, **(19)**].

L'intérêt du Dublin Core réside, entre autres, dans sa simplicité d'implémentation, son extensibilité et flexibilité, son évolutivité et interopérabilité et son multilinguisme [Bureau & Ebacher, 2000, p. 25, **(19)** ; Morel-Pair, 2004, **(15)**]. C'est une norme (ISO 15836-2003) largement plébiscitée. Néanmoins, son expansion est freinée par la réticence des moteurs à indexer ces données.

○ Texte visible de la page

Les mots les plus représentatifs du contenu seront placés haut dans le corps du texte de la page car les moteurs de recherche leur prêtent un indice de pertinence supérieur : « *search engines will also check to see if the search keywords appear near the top of a Web page, such as in the headline or in the first few paragraphs of text. They assume that any page relevant to the topic will mention those words right from the beginning* »⁹ [Sullivan, 2003, **(20)**].

De même, certains moteurs accordent une priorité aux mots mis en exergue (gras, italique, souligné, lien) [Andrieu, 2000, p. 206-207, **(02)**].

⁹ « les moteurs de recherche vérifieront également que les mots-clés apparaissent en haut d'une page Web, dans le titre ou les premiers paragraphes du texte. Ils supposent qu'une page portant sur le thème concerné mentionnera ces mots dès le début. »

○ **Attribut ALT de la balise **

L'attribut ALT de la balise IMG permet de présenter, à la place des images, une description textuelle, prise en compte par les robots. Il faut donc veiller à indiquer, pour chaque image, des mots-clés pertinents par rapport au contenu graphique présenté [Chu, 2003, p. 98, **(03)** ; Andrieu, 2000, p. 137, **(02)**].

○ **Contenu de l'URL**

Lorsque le robot indexe une page web, il tient compte de l'URL. Il se révèle alors judicieux d'inclure les principaux mots-clés dans les adresses des pages, en jouant sur les noms des répertoires et des fichiers [Chu, 2003, p. 100, **(03)** ; Andrieu, 2000, p. 78, **(02)**].

○ **Autres critères**

D'autres critères tels que le poids de la page, la date de mise à jour ou d'entrée dans l'index, sont également susceptibles d'impacter la qualité du référencement, les pages anciennes étant moins bien placées que les plus récentes [Andrieu, 2000, p. 295, **(02)**].

○ **Indice de densité d'un mot-clé**

« *L'indice de densité des mots-clés* est l'un des critères de pertinence les plus importants utilisés par les algorithmes de pertinence* des moteurs de recherche.* » [Andrieu, 2005, p. 3, **(21)**].

Il correspond au nombre d'occurrences du mot demandé sur le nombre total de mots dans la page, sachant que les mots transparents ou vides (le, la, les, l') ne sont pas pris en compte.

$$\text{Indice de densité (ID) d'un mot-clé} = \frac{\text{Nombre d'occurrences du mot-clé dans la page}}{\text{Nombre total de mots dans la page}}$$

Si un terme est présent cinq fois dans une page contenant 100 mots, il sera gratifié d'un ID de 5%. La taille des pages étant susceptible de faire baisser l'ID, il est préférable d'éviter les pages trop longues au bénéfice de pages courtes, centrées sur un seul thème. Alors qu'Olivier Andrieu situe l'indice idéal des mots-clés entre 2,5 % et 5,5 % [Andrieu, 2005, p. 4, **(21)**], pour Zhao il doit être compris entre 6 et 10 % [Zhao, 2004, p. 109, **(22)**], 10 et 18 % pour Saporta [Saporta, 2005, p. 27, **(23)**], Calishain et Dornsfet mettant tout le monde d'accord avec un taux compris entre 5 et 20 % [Calishain & Dornsfet, 2003, p. 327, **(05)**].

Plusieurs utilitaires¹⁰ sont mis à disposition des *webmasters** et des internautes pour calculer l'indice de densité des mots-clés au sein des pages web.

Suite aux tentatives de *spamdexing**, les moteurs ne se contentent plus des critères contenus dans le code HTML des pages (« *in-page* ») pour en évaluer la pertinence mais s'appuient sur des données externes ou « critères *off-page* ».

1.4.2. Critères de pertinence liés au site (« *off-page* »)

Les critères « *off-page* » (présence de mots demandés dans le nom de domaine, indice de popularité* et de clics, etc.), difficilement manipulables, prennent en effet une importance croissante.

○ Nom de domaine

Le nom de domaine peut contenir les mots-clés principaux (ex : immobilier-paris.com). Il est nécessaire de séparer les mots par un tiret car les moteurs ne savent pas dissocier deux mots concaténés [Andrieu, 2005, (17)].

○ Indice de popularité d'une page (IPP)

Il est « *proportionnel au nombre de liens qui mènent d'autres pages web à celle dont l'IPP est calculé.* » [Andrieu, 2000, p. 149, (02)]. Il est égal à l'équation suivante :

$$\text{IPP} = \frac{\text{nombre de pages qui contiennent un lien vers le document}}{\text{nombre de pages total de l'index}}$$

L'IPP est d'autant plus élevé qu'il y a de documents pointant vers une page web. La popularité d'un site est donc tributaire des actions de promotion engagées auprès de sites similaires, en particulier les échanges de liens.

○ Indice de clic

Le moteur prend en compte le comportement de l'internaute (allers-retours sur le moteur jusqu'à satisfaction) pour retrouver les pages les plus « populaires » et reconsidérer leur rang [Andrieu, 2000, p. 152, (02)].

¹⁰ WebrankInfo met un outil de calcul d'indice de densité à la disposition des internautes à l'adresse suivante : <http://www.webrankinfo.com/outils/indice-densite.php> [consulté le 18 août 2005].

1.4.3. Classement des pages (*ranking*)

Le *ranking* définit le classement automatique des données de l'index. Il valorise les sites jugés les plus pertinents en les positionnant favorablement dans la liste des résultats à la suite d'une requête. Pour ce faire, les moteurs se basent sur des algorithmes* souvent secrets voire protégés par des brevets (cf. PageRank©* de Google). On distingue trois méthodes principales de *ranking* [Andrieu, 2005, p. 13, (04)] :

○ Tri par pertinence

Les résultats d'une recherche sont classés selon six facteurs : poids et localisation d'un mot dans le document ou dans son URL, densité d'un mot, mise en relief d'un mot, poids d'un mot dans la base de données, correspondance d'expression entre la requête et l'expression équivalente dans un document, relation de proximité entre les termes de la question et ceux utilisés dans le document.

○ Tri par popularité

Le tri par popularité, laquelle a une influence certaine sur le positionnement, se base sur la co-citation de liens (*backlink* en anglais). Néanmoins, leur nombre n'est pas un critère suffisant pour asseoir une popularité. Le calcul tient donc compte du type (autre outil de recherche, site institutionnel, gouvernemental, etc.) et du statut du site dont le lien émane, de son trafic et de sa place sur la page référente (en haut de page, par exemple). Un site à fort trafic jouit d'une valeur plus importante.

Google ordonne les documents en fonction de leur PageRank© (nombres et qualités des liens pointant vers ces documents). Il tend à favoriser le positionnement des pages issues de sources établies considérant « [...] *qu'une page « pointée » par des sites populaires sera plus pertinente qu'une autre qui réunit le même nombre de liens mais en provenance de sites peu connus.* » [Andrieu, 2000, p. 151, (02)]. Le PageRank© (PR) est exprimé par une valeur comprise entre 1 et 10 : PR entre 1 et 5 : popularité très mauvaise ; PR entre 6 et 7 : popularité moyenne ; PR entre 7 et 8 : bonne popularité ; PR au-delà de 8 : popularité excellente [Saporta, 2005, p. 28, (23)].

La célèbre « Google Dance* » correspond à la période de mise à jour de la base de données de Google et du PageRank© de chacune des pages indexées [WebRankInfo]. A la suite de cette opération, le positionnement d'un site peut varier du tout au tout.

○ Tri par mesure d'audience

Délaissé depuis quelque temps, ce procédé créé par la société DirectHit en 1998, « permet de trier les pages en fonction du nombre et de la « qualité » des visites qu'elles reçoivent. Le moteur analyse alors le comportement des internautes à chaque visite d'un lien depuis la page de résultats pour tenter de trouver les pages les plus « populaires » parmi les pages référencées et améliorer en conséquence leur classement dans les résultats.» [Andrieu, 2005, p. 13, (04)].

○ Tri par catégories

Popularisée en 1997 par Northernlight, cette méthode adoptée notamment par Exalead¹¹ « propose le classement automatique des documents trouvés dans les dossiers et sous-dossiers (clustering*) constitués en fonction des réponses. Les réponses intégrées à chaque dossier sont également triées par pertinence [...]» [Andrieu, 2005, p. 13, (04)].

Les algorithmes changent régulièrement pour contrer le *spamdexing* (voir infra, chapitre 2, p. 48). Les webmestres* doivent donc rester vigilants et suivre de près l'évolution des moteurs pour être capables de réagir rapidement aux nouveaux critères de classement.

¹¹ <http://www.exalead.fr> [consulté le 1^{er} septembre 2005]

1.5. Limites des moteurs de recherche

Malgré des fonctionnalités croissantes et une indéniable efficacité, les moteurs de recherche présentent toutefois quelques lacunes :

- Ils ne peuvent fournir de manière exhaustive l'ensemble des documents réellement disponibles sur Internet, estimés entre 200 et 800 millions¹².
- Parmi les documents référencés, certains sont périmés (informations obsolètes, disparition du site, erreur 404, etc.). En Février 2000, Greg R. Notess répertoriait 13,7 % de liens invalides sur AltaVista et 4,3 % sur Google¹³.
- Les moteurs ne savent pas tout indexer : pages orphelines (non liées à d'autres pages web) ou dynamiques, contenu des bases de données, Flash, Javascript, *frames**, etc. sont encore mal gérés.
- Les méthodes de classement des résultats demeurent opaques et complexes pour la majorité des internautes.
- Il est toujours impossible de trier les résultats de recherche par ordre chronologique ou par auteurs comme le regrettait déjà Rambaux-Yakovleff en 1999 [Rambaux-Yakovleff, 1999, p. 15, (24)].

Ces déficiences n'entachent pas le succès des moteurs de recherche. Les internautes privilégient ce moyen d'accès aux ressources exponentielles du web. Pour ce faire, ils lancent leur requête sur les interfaces dépouillées (Google) ou plus garnies (Yahoo !) de ces outils.

¹² Réseau Education-Médias (2005). *Comment chercher dans Internet de façon efficace ?*.- http://www.media-awareness.ca/francais/ressources/projets_speciaux/toile_ressources/recherche_efficace.cfm [consulté le 23 août 2005]

¹³ NOTESS, Greg R. *Search Engine Statistics : Dead Links Report*.- <http://www.searchengineshowdown.com/stats/dead.shtml> [consulté le 23 août 2005]

2. La recherche avec un moteur

2.1. Moteur d'interrogation

Un moteur effectue sa recherche en texte intégral sur le contenu des pages. Lorsqu'une recherche est lancée, les mots-clés sont comparés aux mots indexés. Puis, le moteur ordonne ces résultats - sous la forme d'une liste de liens - en fonction d'un algorithme de classement qui lui est propre.

Le moteur d'interrogation ou *searcher* est l'interface à partir de laquelle l'utilisateur va soumettre sa question. « *A chaque question, par le biais d'un script CGI (Common Gateway Interface), une requête est générée dans la base de données et une page Web dynamique restitue les résultats généralement sous forme de liste ou de cartes de résultats. L'interface CGI permet d'exécuter un programme sur un serveur et de renvoyer le résultat à un navigateur Internet.* » [Andrieu, 2005, p. 14, (04)].

2.2. Type de recherche

Les moteurs de recherche proposent généralement plusieurs modes d'interrogation :

- la recherche simple : l'internaute soumet des mots et le système retrouve les documents répondant le mieux à la requête dans l'ensemble de son index,
- la recherche avancée avec formulaire de requête : les opérateurs booléens* (AND, OR, NOT) permettent d'affiner la recherche,
- la recherche ciblée : l'internaute peut restreindre sa recherche à un type de documents particuliers (images, textes, MP3/Audio, vidéos) sur des sites nationaux ou sur l'ensemble du web.

L'internaute peut lancer sa recherche sur une expression composée de plusieurs mots, des mots entiers ou des parties de mots. Plus les questions comporteront un grand nombre de termes, plus la réponse sera précise. La plupart des moteurs acceptent la troncature, sont insensibles à la casse et gèrent les fautes d'orthographe [Rambaux-Yakovleff, 1999, p. 11, (24)].

2.3. Page de résultats

Si le poids accordé à chaque critère varie en fonction de l'outil considéré, l'affichage des résultats est relativement homogène. Deux types de contenus se partagent la page : les liens organiques ou « naturels » obtenus grâce à l'indexation automatique des robots, et les liens sponsorisés, payants.

Les réponses s'affichent sous la forme du titre suivi d'un court descriptif de la page (éléments du titre ou premières lignes du texte) ou sous la forme d'une carte ou d'un graphique pour certains moteurs. D'autres, comme Exalead, permettent de reformuler ou de raffiner la question en sélectionnant des termes connexes.

2.4. Sémantique et syntaxe d'interrogation

Les internautes lancent leur recherche sur un mot ou une phrase car ils maîtrisent mal la logique combinatoire et ses opérateurs booléens, complexes pour un néophyte.

Spink, Bateman et Jansen (1999) ont observé que sur 316 utilisateurs d'Excite¹⁴ seuls quelques uns employaient les opérateurs logiques alors qu'il n'était plus qu'une poignée à écrire la syntaxe correctement [Weideman & Strümpfer, 2004, p. 61, (25)].

2.5. Questions générales ou ciblées

De nombreux auteurs se sont intéressés aux stratégies de recherche employées par les internautes. Les données fluctuent en fonction des études.

Certaines d'entre elles prouvent que l'interrogation simple est toujours la plus pratiquée. Nowicki (2003) montre que près de la moitié de ses sujets d'étude n'utilise qu'un seul mot [Zhao, 2004, p. 109, (22)]. Spink, Jansen et Bateman (2001) dévoilent que « *most people use few search terms, few modified queries, view few Web pages, and rarely use advanced search features.*¹⁵ » [Weideman & Strümpfer, 2004, p. 60, (25)]. Enfin, une étude Ad'Oc menée en France sur l'année 2005 révèle que 34 % des parts de recherches concernent la requête en un terme, devant la requête en deux (29,8 %) ou trois (17,8 %) mots [Journal du net, 2005, (26)].

¹⁴ <http://www.excite.com>

¹⁵ « La plupart des gens emploient peu de termes de recherche, peu de questions modifiées, regardent peu de pages, et emploient rarement les dispositifs de recherche avancée. »

D'après Castagné, ce recours instinctif aux questions unitermes s'explique par le principe général « *d'économie cognitive* » qui « *consiste à effectuer des opérations génériques plutôt que des opérations complexes, dans une première phase de recherche* » [Castagné, 2004, p. 27, **(27)**]. Cette stratégie est rarement abandonnée au cours d'une recherche malgré le bruit* ainsi généré.

D'autres études décrivent pourtant une réalité bien différente. Ainsi, Sandrine Saporta, en se basant sur les résultats d'une enquête menée en 2003 par Outrider, juge les requêtes des internautes plutôt ciblées (45 % des recherches comprennent au moins quatre mots et 65 % au moins trois mots) et détaillées puisque 28 % des internautes précisent le nom d'un produit et 9 % celui d'une marque [Saporta, 2005, p. 46-47, **(23)**]. Selon Référenceur.com, les internautes utilisent en moyenne 2,5 mots par requête pour atténuer le bruit engendré par la recherche uniterme [Referenceur.com, **(28)**]. Enfin, sur les 540 utilisateurs suivis par Weideman & Strümpfer [Weideman & Strümpfer, 2004, p. 62, **(25)**], seuls 109 utilisent un unique mot-clé. Les autres (431 personnes) en tapent davantage.

Aujourd'hui, la plupart des portails de recherche mondiaux ou francophones offrent à la fois un annuaire proposant des fiches descriptives de sites (de type Yahoo! Directory) et un moteur de recherche en texte intégral sur les pages du web (de type Google) [abondance.com].

La présence d'un site web sur ces outils est essentielle pour assurer sa visibilité. Elle va dépendre de l'inscription du site sur les moteurs et annuaires, une démarche connue sous le nom de « référencement ».

CHAPITRE 2

Le référencement : une technique documentaire ou marketing ?

1. Intérêts et enjeux du référencement

1.1. Définition

Au sens strict, le référencement définit l'action d'enregistrer un site web sur un outil de recherche, afin de le faire connaître aux internautes. « *Dans son acception large, [il] peut être défini comme l'ensemble des actions permettant à un site d'être présent dans les bases de données des outils de recherche* » [Saporta, 2005, p. 19, (23)]. Il poursuit un double objectif : « *d'une part inscrire le site Internet dans les bases de données des outils de recherche ; d'autre part, le positionner parmi les premières pages de résultat.* » [Saporta, 2005, p. 17, (23)].

1.2. De l'importance d'être référencé sur les moteurs de recherche

Selon une étude de décembre 2004 publiée dans la Revue du Référencement, les moteurs de recherche génèrent 42,67 % des visites sur un site, les liens externes 26,14 % et les accès directs 31,19 % [Revue-Référencement, 2005, (29)]. Si, d'autre part, on considère que quatre fois sur dix un e-consommateur visite un site marchand grâce à un moteur de recherche (étude Overture 2003) [Saporta, 2005, p. 9, (23)], on mesure l'importance pour celui-ci d'être référencé sur ce type d'outils.

1.3. Du référencement au positionnement

L'intérêt du référencement est de générer un trafic ciblé. Mais être référencé auprès des principaux outils de recherche ne suffit plus. Vu la concurrence présente sur la Toile, le positionnement sur des mots-clés précis s'avère désormais fondamental [Chu, 2003, p. 95, (03)]. Le positionnement a pour objectif de placer un site parmi les premières pages de résultat afin de lui assurer un maximum de visibilité. Idéalement, un site gagne à être positionné dans les premières places. En 2000, une enquête sur les utilisateurs d'Excite a montré que 77 % des internautes ne s'aventuraient pas au-delà des deux premières pages de résultat. Une autre étude menée en 2003 dévoilait pour sa part que 85 % des utilisateurs des moteurs de recherche ne s'intéressaient pas aux résultats classés après la 40^e place [Xue, 2004, p. 188, (30)].

1.4. Différence entre indexation et référencement

Comme nous l'avons vu précédemment (voir supra « Différence entre indexation documentaire et indexation informatique, p. 20), le but premier de l'indexation, quel que soit le système utilisé - manuel ou automatisé -, est de faciliter l'accès rapide au contenu d'un ou plusieurs documents. C'est une démarche essentielle au service de la recherche documentaire. Quant au référencement, il consiste à soumettre l'adresse des sites web sur les outils de recherche. C'est avant tout un acte promotionnel, destiné à rendre visible un document indexé.

Sur le Web, ces deux process sont fortement imbriqués. En effet, deux cas de figure sont envisageables :

- le référencement du site sur un outil va déclencher son indexation,
- l'indexation automatique par les robots va aboutir au référencement du site.

Finalement le référencement n'est rien d'autre que la déclaration officielle de naissance du site.

Indexation documentaire, indexation informatique et référencement désignent des réalités différentes qu'il ne faut pas confondre. Rappelons que la première de ces notions définit le « processus destiné à représenter [...] les notions caractéristiques du contenu d'un document [...] en vue d'en faciliter la recherche, après les avoir identifiées par l'analyse » [Boulogne, 2004, p. 123, **(31)**]. La seconde décrit l'action d'un robot repérant les documents sur le web et les analysant (syntaxe, liens, etc.) pour les ajouter à l'index du moteur. Enfin, le référencement désigne l'« opération qui consiste à répertorier l'adresse des sites web sur des annuaires ou des moteurs de recherche afin d'en faciliter l'accès et de développer le trafic sur ces sites. » [Cacaly et al. , 2004, p. 193, **(07)**].

Une fois inscrit sur les outils de recherche, le site devient visible aux yeux des internautes et apte à générer du trafic pour peu qu'il ait été préalablement optimisé.

2. Optimiser une page Web pour le référencement

La première mission du référencement est d'assurer la présence d'un nombre maximal de pages d'un site sur les outils de recherche.

2.1. Les règles générales de l'optimisation

Un bon positionnement dépend de l'adéquation du code source de la page Web avec les critères de pertinence propres à chaque moteur, lesquels sont connus et faciles à mettre en œuvre (voir supra Chapitre 1, « Méthodes de tri des résultats et *ranking* », p. 22). Il est dès lors surprenant de constater que seuls 7 % des sites commerciaux sont optimisés pour un référencement [Saporta, 2005, p. 32, **(23)**].

Outre les préconisations déjà évoquées au Chapitre 1, rappelons que les algorithmes de classement des moteurs de recherche attribuent une pondération supérieure aux 200 premiers caractères d'une page web. Il est donc recommandé d'éviter les pages totalement graphiques [Chu, 2003, p. 97, **(03)**] et d'utiliser autant que possible les mots dans les titres, les liens, les attributs ALT des images et dans le texte.

Mis à part la balise « meta robots », qui permet d'interdire l'indexation d'une page par les moteurs de recherche et de suivre ou non les liens qu'elle contient, les autres balises meta n'ont plus aucune incidence dans le cadre d'un référencement [Andrieu, 2005, **(17)**].

Enfin, l'emplacement de la page dans l'arborescence d'un site a également un impact direct sur son référencement : les robots accèdent plus difficilement à une page située dans un sous-répertoire de 5^e niveau qu'à une page située à la racine d'un site. En règle générale, il est conseillé de ne pas dépasser plus de trois niveaux d'arborescence [Chu, 2003, p. 100, **(03)**].

Les démarches de référencement gagnent à être effectuées une fois que le site est finalisé (pas de liens inactifs ni de pages en construction).

2.2. Sélection des mots-clés

Le besoin d'information est un processus cognitif individuel, propre à un utilisateur. Il va influencer sa recherche et sa lecture du document obtenu. Dans ces conditions, comment déterminer la nature des mots-clés que va utiliser l'internaute pour chercher l'information ?

Le choix des mots-clés (mots isolés ou expressions regroupant plusieurs mots) est une étape fondamentale dans le processus de référencement d'un site web. En effet, « *chaque combinaison ou mot aura une incidence sur l'optimisation et donc le placement du site et des pages web sur les outils de recherche* » [Chu, 2003, p. 98, **(03)**]. Cette problématique est familière aux documentalistes qui s'interrogent depuis longtemps sur la détermination des termes représentatifs d'un document, leur emploi effectif par les utilisateurs lors de leurs recherches et les moyens d'atténuer le bruit et le silence* des résultats.

2.2.1. Définition

En informatique, le concept de « mot-clé » diverge de son sens documentaire. En effet, il désigne un mot non normalisé, issu du vocabulaire libre, qui se réfère non à l'attribut « Keywords » comme on le croit souvent, mais aux mots du texte. C'est pourquoi, il semble plus judicieux d'employer le terme « descripteur* » ou « vedette-matière » dans le cadre strictement documentaire afin de bien les différencier.

On peut distinguer deux types de mots-clés [Andrieu, 2001, p. 68, **(11)**] : les mots primaires (ou « larges ») qui définissent le domaine d'intervention (agroalimentaire, informatique, presse, histoire, santé, médecine, etc.) et les mots secondaires (ou « profonds ») plus précis, qui permettent d'approfondir la recherche en fonction des besoins de l'internaute.

2.2.2. Choix des mots-clés

Sur le web, les mots-clés les plus rentables sont ceux qui génèrent du trafic. Deux objectifs doivent donc guider la définition de ces mots : ils doivent refléter le contenu du site et s'adapter à la cible visée. Plus elle est réduite, plus le vocabulaire utilisé sera précis et la liste de mots-clés clairement délimitée. En revanche, si le site concerne un large public, notamment dans le cadre d'un objectif de notoriété, le vocabulaire sera plus général et donc plus approximatif [Saporta, 2005, p. 55, **(23)** ; Chu, 2003, p. 98, **(03)**].

Le moyen le plus efficace de choisir les bons mots-clés consiste à se substituer aux internautes effectuant une recherche et à faire preuve d'intuition pour envisager toutes les requêtes susceptibles d'être posées [Chirié, 1995, p. 65, **(10)** ; Calishain & Dornsfet, 2003, p. 340, **(05)** ; Quatravaux, 2004, p. 249, **(32)**].

On peut également s'inspirer des indications offertes par les sociétés qui vendent du positionnement publicitaire (AdWords*, Overture) ou encore des nombreux utilitaires disponibles sur Internet - ils permettent d'obtenir, pour un mot-clé, les combinaisons complémentaires pour ce mot ainsi que le nombre de recherches associées¹⁶ - dont le site Abondance.com fournit une liste non exhaustive¹⁷.

Dans leur ensemble, les auteurs [Le Guelvouit, 1999, p. 129, **(33)** ; Saporta, 2005, p. 55 et 73, **(23)**] s'accordent à souligner l'importance, pour une page ou un site Web, de se positionner sur les mots-clés populaires ou familiers et les abréviations, synonymes et anglicismes. Le référenceur doit également profiter des erreurs de frappe et fautes d'orthographe pour se positionner au mieux.

Alors que, pour l'indexation, l'emploi du singulier est recommandé par une norme (NF Z 44-070), dans le cadre d'un référencement, au contraire, il faut prévoir toutes les variantes orthographiques d'un mot (pluriel, singulier, lettres accentuées ou non).

L'indexation vise à faciliter l'accès rapide à l'information. Le référencement a, en sus, une visée marketing. Toutes les variantes d'un même mot, qu'il soit populaire ou mal orthographié, seront clés de recherche pour l'internaute, ce qui est impensable dans le cadre de l'indexation documentaire.

¹⁶ http://www.indicateur.com/Doc/generateurs_mots_cles.shtml [consulté le 31 août 2005]

¹⁷ <http://ressources.abondance.com/generateur-mot-cle.html> [consulté le 29 août 2005]

2.3. Facteurs bloquants

De nombreux facteurs peuvent influencer sur le référencement d'un site. Le cahier des charges, préalable à toute création, devra prendre en compte ces contraintes.

2.3.1. Les *frames*

Il s'agit d'un système qui permet d'afficher plusieurs pages indépendantes dans une même fenêtre du navigateur. En général, on utilise les *frames* pour afficher une zone de menu statique et répétitive alors que le reste de la page est visible dans une zone que l'on fait défiler. La balise <frameset> n'est pas reconnue par tous les moteurs. Du coup, les robots ne naviguent pas correctement dans le site et n'indexent pas toutes les pages. L'insertion de la balise <noframes> peut toutefois permettre de résoudre ce désagrément [Chu, 2003, p. 102, (03)].

2.3.2. Flash et Javascript

Flash et Javascript nuisent à l'indexation car ils ne sont pas gérés correctement par les robots. Les informations présentes dans une animation Flash sont ignorées, les liens en JavaScript (pop-up, rollover...) non suivis. Il est préférable de restreindre leur utilisation, de proposer une version sans Flash du site et de doubler les liens graphiques par une version textuelle simple [Chu, 2003, p. 103, (03)].

2.3.3. Les pages dynamiques

Certains moteurs rencontrent des difficultés à indexer les pages dynamiques (voir infra « Le référencement des pages dynamiques », p. 46). Ces dernières sont identifiables à leur URL « exotique »* comprenant des points d'interrogation (?) et esperluettes (&) à l'image de l'adresse ci-dessous :

<http://www.lavoisier.fr/fr/livres/index.asp?texte=r%E9f%E9rencement&select=motcle&exact=on&togo=&support=NULL&from=>

Parmi les autres facteurs bloquants, citons en vrac : le script CGI en page d'accueil, les pages trop graphiques, les balises <meta refresh> ou les redirections réalisées au niveau du DNS (Domain Name Server) et le *spamdexing* [Andrieu, 2000, p. 248-249, (02)].

3. Un référencement différencié en fonction des outils de recherche et des objectifs

Les techniques de référencement varient non seulement en fonction des outils de recherche mais aussi des stratégies de visibilité (services gratuits ou payants).

3.1. Référencement auprès des annuaires

Pour apparaître dans un annuaire, le référencement est obligatoire. L'inscription passe par le remplissage volontaire d'un formulaire.

Avant de soumettre un site à un annuaire, il est impératif de déterminer la catégorie la plus en adéquation avec le thème du site. Plusieurs paramètres sont à prendre en compte dont les niveaux d'arborescence (le site doit être facilement accessible), la visibilité du site au sein de la catégorie (choisir des rubriques qui ne sont pas surchargées), la présence de concurrents [Chu, 2003, p. 104, **(03)** ; Andrieu, 2000, p. 350-351, **(02)**]. Une fois la catégorie définie, on remplit une fiche d'information. Elle est généralement composée de deux parties : la carte d'identité du site (titre, URL, description, situation géographique, etc.) et des informations sur le contact.

Il faut encore patienter quelques semaines voire des mois avant que l'équipe éditoriale ne prenne en compte la soumission*. Cette dernière ne garantit pas l'inscription. L'acceptation définitive du site dépend des netsurfeurs [Saporta, 2005, p. 18, **(23)** ; Foenix-Riou, 2002, p. 10, **(34)**]. S'il est toujours possible de soumettre un site gratuitement, de nombreux annuaires rendent l'inscription obligatoire pour les sites de e-commerce (Yahoo !). L'offre de soumission gratuite ne garantit aucunement la visite du netsurfeur sur le site à la différence de l'offre de soumission payante* qui l'assure dans un délai clairement spécifié.

Visiblement, les référenceurs apprécient le service payant. Il leur permet d'apporter des garanties à leur clientèle [Foenix-Riou, 2002, p. 11, **(34)**].

3.2. Référencement classique auprès des moteurs de recherche

3.2.1. Référencement naturel

Le référencement naturel définit toutes les techniques de référencement gratuit. René Cohen, PDG de Ciblo.net reproche le manque de fiabilité et la piètre performance de cette technique de référencement « [...] *car il est très difficile de faire remonter un site parmi les premiers résultats de recherche sur une sélection définie de mots clés.* » [Saporta, 2005, p. 103, (23)], d'autant que la visibilité des pages ainsi référencées pâtit de l'omniprésence des liens sponsorisés dans les premières pages de résultats.

3.2.2. Référencement par indexation

Elle permet au site d'être traité en priorité (généralement une semaine) et certifie la présence d'un certain nombre de pages dans l'index du moteur ainsi que le rafraîchissement des documents selon une périodicité donnée [Saporta, 2005, p. 19, (23) ; Foenix-Riou, 2002, p. 10, (34)]. En revanche, elle ne garantit pas le positionnement.

3.2.3. Référencement par positionnement

« *Positionner un site signifie tenter de placer, en réponse à un mot ou une requête, une page d'un site dans les X premiers résultats des outils de recherche.* » [Saporta, 2005, p. 22, (23)].

Le développement de cette véritable démarche marketing s'appuie sur l'analyse d'une constante repérée par plusieurs études sur le comportement des utilisateurs d'outils de recherche. D'après elles, les internautes focalisent leur attention sur les premières réponses [Zhang & Dimitroff, 2004, p. 310, (35)] et particulièrement sur les deux premières pages de résultats [Le Guelvouit, 1998, p. 316, (36) ; Saporta, 2005, p. 17, (23)]. Un site placé en première position obtient en moyenne trois fois plus de clics que le cinquième [Saporta, 2005, p. 78, (23)]. Etre classé dans les dix premiers résultats proposés par les moteurs, ou au pis dans les trente premiers sur les mots-clés essentiels [Andrieu, 2000, p. 145, (02)] s'avère donc avantageux.

Toutefois, le positionnement ne peut se faire que sur des mots ou expressions bien ciblés. Le nombre de résultats est un bon indicateur de la faisabilité d'un référencement concurrentiel. Ce dernier s'annonce particulièrement difficile si la recherche sur l'expression désirée génère plus de 20 000 réponses [Andrieu, 2000, p. 394, (02)].

Jamais acquis et en perpétuelle évolution en raison des variations des modes de calcul et critères de classement des moteurs, un référencement par positionnement se construit sur du long terme et exige un suivi rigoureux et des corrections régulières [Saporta, 2005, p. 23, **(23)**].

Ces contraintes expliquent le développement du positionnement payant. Pour un mot-clé donné, l'éditeur est assuré de la présence de ses pages ou de son site dans les premiers résultats du moteur [Foenix-Riou, 2002, p. 10, **(34)**].

A l'occasion d'une journée consacrée au référencement payant organisée le 28 mai 2002, les sociétés spécialisées soulignaient ses avantages (facilité de gestion des campagnes événementielles, positionnement de sites sur les mots-clés les plus utilisés, risques faibles) mais aussi ses inconvénients (coût très important, chronophagie) [Foenix-Riou, 2002, p. 12, **(34)**].

3.3. Quand le référencement est devenu de la publicité...

Le référencement payant est né à la fin de l'année 2000 aux Etats-Unis. Il est apparu en France en 2001. Depuis, il s'est généralisé à l'ensemble des moteurs de recherche [Foenix-Riou, 2002, p. 10, **(34)**]. Il apporte l'assurance aux éditeurs que leurs sites seront bien présents dans l'outil de recherche sélectionné.

Ces outils publicitaires mobilisent quatre intervenants : un annonceur* (qui paye pour diffuser sa publicité), un éditeur (qui accueille la publicité sur son site contre une rétribution), une régie (qui comptabilise les affichages et les clics et facture l'annonceur en conséquence) et un internaute [Carlier, 2005, **(37)**].

Trois régies (Miva ex-Espotting, Overture et Google) centralisent l'achat des espaces publicitaires sur les outils de recherche. Un annonceur est contraint de passer par ces régies pour réaliser sa campagne de communication [Saporta, 2005, p. 122, **(23)**].

3.3.1. Liens sponsorisés

Le lien sponsorisé se définit comme un support publicitaire en ligne « *qui permet de référencer un site Internet parmi les premières positions de résultats des recherches [...]* » [Saporta, 2005, p. 44, **(23)**]. Il « *est une réponse (= offre d'un annonceur) à une requête formulée par un internaute (= demande d'un prospect) [...]* » [Saporta, 2005, p. 147, **(23)**]. En France, les liens sponsorisés (MSN, AltaVista), appelés également « liens commerciaux » (Google) ou « liens promotionnels » (Yahoo) selon les moteurs, représenteraient 30 % du marché publicitaire sur Internet [Meslon, 2005, **(38)**].

Ils arrivent tous en haut de page et sont généralement limités à trois ou quatre sur un même mot-clé. C'est un moyen de communication efficace et meilleur marché que les méthodes marketing traditionnelles, qui, de plus, assure une visibilité optimale aux annonceurs.

La « facturation à la performance », largement développée, est basée sur un système d'enchères : la première place dans les résultats est dévolue à l'annonceur proposant le prix le plus élevé par clic sur un mot en rapport avec sa thématique [Saporta, 2005, p. 48, **(23)**]. La saisie de ces mots dans le moteur déclenchera l'affichage de bandeaux ciblés sur la page d'un éditeur. Selon le mode de facturation au coût par clic* (CPC), l'annonceur ne paiera que lorsqu'un internaute cliquera sur un lien sponsorisé.

Une campagne de liens promotionnels peut répondre à différents objectifs : créer du trafic sur un site Internet pour maximiser un retour sur investissement*, développer la notoriété d'une marque, acquérir une base de données de prospects qualifiés ou de nouveaux clients, prendre des parts de marché à la concurrence, appuyer une action de communication *off-line* ou animer des opérations ponctuelles [Saporta, 2005, p. 45-46, **(23)**]. Sa mise en place nécessite donc des connaissances particulières en Web marketing afin de calculer le retour sur investissement, le coût d'acquisition* client, d'anticiper et de réagir aux attitudes des internautes en phase de recherche d'information [Saporta, 2005, p. 49, **(23)**].

Le choix de mots-clés finement ciblés et qualifiés joue sur les retombées d'une campagne de liens sponsorisés. Leur sélection est fondamentale dans le cadre d'un objectif de retour sur investissement.

En revanche, dans le cadre d'un objectif de visibilité et de notoriété, l'annonceur veillera à être présent sur le plus grand nombre de mots-clés possibles et surtout sur les mots-clés génériques pour générer un trafic conséquent [Saporta, 2005, p. 54 et 70, **(23)**].

La nature du positionnement (première place impérative ou simple présence) est à déterminer en fonction de chaque mot-clé. Les plus génériques (logiciel, livre) gagneront à être accompagnés de descriptifs précis pour renseigner au mieux l'internaute et garantir de fait un trafic qualifié*. Toutefois, à moins d'adopter une stratégie de positionnement pour affirmer un *leadership*, il vaut mieux éviter de se positionner en première position sur ce type de mots en raison de leur coût élevé [Saporta, 2005, p. 77, **(23)**].

La réussite d'une campagne de liens sponsorisés se mesure moins au nombre de clics générés qu'au taux de transformation*. Ce dernier correspond au pourcentage de visiteurs qui transforment leur visite en acte d'achat, d'inscription ou d'abonnement [Saporta, 2005, p. 61, **(23)**].

Si Yahoo!¹⁸ ou Google¹⁹ affichent clairement les liens sponsorisés dans un encadré bleuté au-dessus de la liste des sites référencés naturellement ou sur une colonne à droite, sur Nomade²⁰, la présentation n'est pas claire : seule la notion « sponsorisé par » (à la place de l'indication « adresse »), signale qu'il s'agit d'un lien promotionnel.

3.3.2. Liens contextuels

« Les liens contextuels sont des liens publicitaires textuels présents sur des pages de contenu [...] » [Saporta, 2005, p. 83, **(23)**]. Ce mode de communication crée du trafic qualifié. En effet, les liens contextuels s'affichent en fonction du contenu de la page et non d'une requête [Meslon, 2005, **(38)**].

Google AdSense est le programme de liens contextuels mis en place par Google. Il allie la publicité au coût par clic (CPC) et la publicité au coût pour mille impressions* (CPM), mis en concurrence au sein d'un même système d'enchères.

¹⁸ <http://fr.yahoo.com>

¹⁹ <http://www.google.fr>

²⁰ <http://www.nomade.tiscali.fr>

Les annonceurs paient soit lorsqu'un utilisateur clique sur une annonce, soit lorsque leur annonce s'affiche sur un site. De son côté, l'éditeur reçoit une rémunération versée par Google en fonction des clics effectués sur les annonces.²¹

3.3.3. Pay Inclusion

Le *Pay Inclusion* regroupe deux types d'offres dont la première tombe peu à peu en désuétude :

○ L'inclusion payante d'URL

Elle garantit la visite régulière du robot d'un moteur de recherche moyennant le paiement d'un coût fixe et éventuellement d'un coût par clic. Les résultats ne sont pas signalés comme des résultats sponsorisés. Si elle assure l'indexation ou le rafraîchissement régulier (tous les un ou deux jours) des pages Web d'un site dans l'index, elle n'offre en revanche aucune garantie de positionnement ou de visibilité.

○ Le flux XML ou *Trust Feed*

Le « flux XML » appelé aussi *trust feed*, *pay for inclusion* (PFI), « XML PI » (XML « paid inclusion »), *paid inclusion* ou « Inclusion dynamique » est proposé par les moteurs de recherche depuis mi-2003 en France. Il « permet, par transfert de flux, d'indexer une base de données sur les moteurs de recherche, avec une actualisation fréquente. » [Andrieu, 2005, p. 13, (39)]. Il concerne spécifiquement les sites dynamiques (développés en PHP, ASP, CIG ou CFM) et leurs bases de données, difficiles à indexer de manière traditionnelle, qu'il permet de référencer via le format XML, sans intervention technique lourde [Saporta, 2005, p. 86, (23)]. Cette solution leur permet d'être mieux positionné sur les pages de résultat. Elle est donc particulièrement adaptée au référencement des catalogues produits des sites de e-commerce qui actualisent fréquemment leur base de données [Saporta, 2005, p. 89, (23) ; Andrieu, 2005, p. 17, (39)]. Les moteurs sont rémunérés en fonction du nombre de clics (offre Pay Per Clic).

²¹ <https://www.google.com/adsense/?sourceid=ASO&subid=FR-ET-ABOUT-FR&hl=fr> [consulté le 23 août 2005]

Les professionnels du référencement apprécient la garantie et la rapidité d'indexation, la mise à jour régulière et le coût modique du service [Andrieu, 2005, p. 20-21, **(40)**]. Parmi les inconvénients, ils regrettent les délais de mise à jour, la limitation dans le choix des mots-clés, les problèmes de clics non qualifiés et pourtant facturés par les moteurs, l'optimisation impérative des pages avant soumission, l'évaluation difficile du retour sur investissement et les problèmes de pérennité de l'offre (Google, qui représente la part majoritaire du trafic Internet, n'en propose pas) [Andrieu, 2005, p. 12, 17-18, **(39)**]. Enfin, l'opacité des circuits d'achats pour l'annonceur et du mode de diffusion pour l'internaute, le rapport coût/apport trafic médiocre, l'absence de garantie de positionnement, et le risque de pollution des résultats naturels [Andrieu, 2005, p. 21-23, **(40)**] nuisent à son développement.

Les revenus générés par les liens sponsorisés et le *paid inclusion* s'élèvent à 2,6 milliards de dollars en 2004. Ces revenus se répartissent comme suit : 4 % proviennent du *paid inclusion* et 96 % des liens sponsorisés [Andrieu, 2005, p. 11, **(39)**].

3.4. Le référencement des pages dynamiques

« Par opposition à une page statique qui est toujours affichée de la même façon (celle qu'a écrite le webmaster), une page dynamique* affiche des informations différentes selon le contexte, en puisant des données dans une base de données. » [WebRankInfo].

Ces pages se reconnaissent à leurs URL complexes contenant en général un point d'interrogation, éventuellement une esperluette (&), ainsi que des noms de variables et des valeurs. Si certains moteurs (Google) indexent ce type de page, la plupart des robots écartent ces URL dites « exotiques » dans le souci de combattre le *spamdexing* ou d'éviter des doublons d'entrées dans leurs bases de données.

Des solutions telles que la création de pages uniquement dévolues au référencement (pages satellites* ou alias) ou la création de *framesets* se révèlent toutefois moins performantes que la réécriture d'URL (*URL rewriting*). Celle-ci permet de convertir les adresses des pages dynamiques en URL apparemment statiques grâce à un paramétrage particulier du serveur web. Cette technique est particulièrement intéressante pour des sites volumineux et complexes [Adcom, 2003, **(41)**].

3.5. Référencement manuel ou référencement automatique ?

Deux méthodes permettent de référencer un site sur les outils de recherche :

- le « référencement manuel » : il se pratique sur chaque outil de recherche individuellement, ce qui permet de prendre en compte leurs particularités et de choisir très précisément l'outil sur lequel on souhaite apparaître. Cette démarche est particulièrement chronophage [Le Guelvouit, 1999, p. 127, **(33)**].

- le référencement automatique* sur plusieurs outils de recherche se fait simultanément via un unique formulaire. Cette option permet de gagner du temps et de simplifier la démarche mais peut néanmoins dégrader la qualité du référencement [Le Guelvouit, 1999, p. 128, **(33)**].

Le référencement automatique peut être effectué par le biais d'un site web spécialisé dans ce domaine ou par l'intermédiaire d'un logiciel dédié [Andrieu, 2000, p. 341, **(02)**]. S'il n'est pas indiqué d'utiliser ce type d'outil, mal adapté aux annuaires (choix d'une catégorie), et rejeté par certains moteurs, il peut s'avérer intéressant pour étendre le référencement à des outils de recherche spécialisés [Chu, 2003, p. 108, **(03)**].

« Le référencement manuel va traiter le côté qualitatif du référencement du site, tandis que le référencement automatique va s'occuper du quantitatif. » [Andrieu, 2000, p. 341, **(02)**].

4. Dérives du référencement

4.1. La vague du payant

La généralisation des offres payantes de référencement s'est affirmée depuis 2001 et d'après Sandrine Saporta, le recours à ces services constitue dorénavant un élément essentiel d'une stratégie réussie de référencement [Saporta, 2005, p. 21, **(23)**].

Si la plupart des outils de recherche continuent à proposer des formules de soumission gratuite, d'autres, en revanche, sont passés au tout payant. Olivier Andrieu regrette la complexité grandissante de rentrer dans un annuaire par le biais d'une soumission gratuite.²²

Le prix se justifie par un délai garanti de référencement (moins d'une semaine), une réactualisation fréquente et régulière des pages ou un positionnement concurrentiel sous forme de « liens sponsorisés » sur une page de résultat pour des mots-clés spécifiques (voir supra : « Liens sponsorisés », p. 43-44) [Chu, 2003, p. 107, **(03)**].

4.2. Le *spamdexing*

Les enjeux commerciaux liés au positionnement des sites sur les moteurs sont à l'origine du *spamdexing* – terme formé par la contraction de *spam* et *indexing* -. Ce concept désigne toutes les techniques destinées à tromper illégalement les algorithmes des robots pour améliorer artificiellement la position d'un site web sur un moteur de recherche (répétition abusive des mêmes mots-clés ou des titres, texte invisible, usurpation de nom ou de marque connus, certaines techniques de *cloaking**, plagiat, pages satellites ou alias*, mise en place de faux liens pour augmenter la popularité, soumissions multiples) [Andrieu, 2000, p. 154, 158-179, **(02)**].

²² Abondance : Les soumissions gratuites sont-elles encore prises en compte par les annuaires ? (Septembre 2002 – Révision mars 2005), <http://docsabondance.com/question52.html> [consulté le 15 août 2005]

En 1998, Le Guelvouit [Le Guelvouit, 1998, p. 76, **(36)**] s'alarmait de « (...) *l'anarchie la plus totale* » régnant sur la Toile en ce domaine. Depuis, les moteurs ont développé des techniques pour repérer ces pratiques. Elles aboutissent à des pénalités (page ignorée et non indexée, mauvais classement), voire au placement automatique des sites qui en abusent sur une liste noire, ce qui bloque toute possibilité de ré-indexation ultérieure sur le moteur [Saporta, 2005, p. 169-170, **(23)** ; Chu, 2003, p. 114, **(03)** ; Andrieu, 2000, p. 155, **(02)**]. En effet, ces pratiques anticoncurrentielles nuisent à la pertinence des résultats, et par conséquent à la réputation du moteur.

4.3. Le positionnement : une perversion de l'indexation ?

Le recours au positionnement offert par des sociétés spécialisées (voir infra : « Les sociétés spécialisées dans le référencement », p. 50) n'est pas sans risque. En effet, ces entreprises peuvent tenter de vendre les premières positions sur des expressions qui flattent le client mais ne sont pas rentables au final [Andrieu, 2000, p. 396, **(02)**].

Une autre dérive couramment constatée concerne l'achat de mots-clés* de marques déposées par des concurrents. Google et Overture ont déjà été condamné suite à de tels abus [Meslon, 2005, **(38)**].

Dans le même registre, certains n'hésitent pas à augmenter artificiellement les enchères sur les mots-clés des marques concurrentes. Ils contraignent ainsi leurs adversaires à payer plus cher pour bénéficier d'un positionnement compétitif [Carlier, 2005, **(37)**].

5. Les sociétés spécialisées dans le référencement

Deux solutions sont envisageables pour assurer le référencement et le positionnement d'un site web : le traitement en propre ou l'externalisation / sous-traitance à un professionnel.

L'actualité des outils de recherche évoluant tous les jours, faire appel à une société spécialisée peut s'avérer rentable car elle suit de très près les évolutions du secteur tant d'un point de vue technique que marketing. Le travail du bon référenceur est de conseiller son client quant aux mots-clés ou expressions sur lesquels il est techniquement possible, d'une part, de faire un positionnement, d'autre part, de générer du trafic, ce qui devient de plus en plus complexe [Andrieu, 2000, p. 395, **(02)**].

Outre le pur référencement, les agences rebaptisées pour l'occasion « agence de *search engine marketing* » ou « agence de *web marketing* », afin de refléter leur évolution, gèrent désormais les campagnes de liens promotionnels [Quin, 2004, **(42)**].

Les relations entre le référenceur prestataire de service et son client s'inscrivent dans le cadre d'un contrat de référencement [Chu, 2003, p. 108, **(03)**]. L'agence confère des garanties aux clients et à la concurrence quant aux méthodes employées, au suivi, à la veille et au respect des critères d'indexation [Saporta, 2005, p. 32, **(23)**].

Ce métier conserve néanmoins un caractère artisanal. Les professionnels doivent continuellement se tenir informés et procéder à des tests pour comprendre au mieux les versatilités des algorithmes [Saporta, 2005, p. 34, **(23)**] et réajuster leur méthodologie en conséquence. Aucune réelle garantie n'est envisageable sur un marché où seuls les moteurs maîtrisent la technologie [Andrieu, 2005, p. 29, **(43)**].

La promotion réussie d'un site web est tributaire d'un bon référencement, capable de lui assurer une visibilité optimale. Il est toutefois important de garder à l'esprit l'évolution continue des technologies basées sur des calculs et des algorithmes souvent secrets. L'essor considérable du marché des liens publicitaires repousse parfois les résultats du référencement naturel à des positions très éloignées (cf. Voila.fr).

CHAPITRE 3

Le référencement : une étape dans le plan marketing d'un site web

1. Le site web : un produit

« *Le marketing est l'ensemble des actions, outils et techniques mis en œuvre en vue de promouvoir la croissance d'une activité.* » [Chu, 2003, p. 131, **(03)**].

Le raisonnement marketing s'applique à tout service s'adressant à un public [Muet & Salaün, 2001, p. 14 **(44)**]. La notion de produit définit donc tout autant « *un bien matériel ou immatériel, un objet, un service, une activité, une personne, une idée, un endroit, une organisation, capables de satisfaire un service ou un besoin.* » [Bernadet et al., 2001, p. 4, **(45)**]. L'approche marketing se base sur le principe que la mission de toute organisation est d'adapter l'offre (un produit, un service) à la demande du client (consommateur ou entreprise).

Le référencement est un moyen de communication qui permet d'atteindre plusieurs objectifs commerciaux : développer les ventes et actions d'un annonceur, assurer un trafic qualifié constant sur un site Web, prendre des parts de visibilité et des clients à la concurrence et développer la fidélisation des internautes [Saporta, 2005, p. 5, **(23)**].

Sandrine Saporta compare le référencement sur Internet à une stratégie « pull ». Dans la mesure où l'internaute tape sa demande (ex : « promo billet avion Montréal »), il fait la démarche d'aller à la recherche du produit [Saporta, 2005, p. 7, **(23)**]. Le positionnement stratégique d'un annonceur sur une requête précise d'un internaute représente une opportunité de rencontrer des prospects qualifiés, au moment de leur intention d'achat.

Le ciblage s'effectue par l'internaute lui-même lorsqu'il tape les termes de sa requête. D'où l'importance de choisir avec soin les mots-clés sur lesquels l'annonceur souhaite être bien positionné puisque ce ciblage précis et qualifié permettra de rentabiliser l'opération en réduisant les coûts d'acquisition clients. Un bon positionnement renforcera également l'image de *leader* des grandes marques [Saporta, 2005, p. 8, **(23)**].

2. Les préalables au référencement

L'analyse de l'environnement et du public, préalable nécessaire à la mise en place d'une démarche stratégique plus générale, est destinée à cerner les objectifs de cette dernière.

2.1. Analyse de l'environnement

Il est nécessaire de connaître l'amont (les moteurs de recherche) et l'aval (les internautes et leurs stratégies de recherche). Une étude du secteur d'activité se révèle également nécessaire. Elle s'intéressera plus particulièrement aux points suivants : détermination de la cible, des activités et des spécificités de l'annonceur, choix des rubriques sur un annuaire, étendue et nombre de produits proposés. Ces données vont agir non seulement sur le choix des mots-clés mais aussi sur la stratégie d'enchères des liens sponsorisés [Saporta, 2005, p. 53, **(23)**].

2.1.1. Analyse des sites concurrents

Le positionnement permet d'affirmer une identité face aux autres offres existantes. La première analyse à mener concerne les acteurs en présence. Il s'agit de connaître leurs produits et de déterminer leurs parts de marché, leurs forces et faiblesses, leur stratégie... Le positionnement d'un site est tributaire de l'intérêt qu'il éveille chez d'autres sites, lesquels pointeront vers lui s'ils l'estiment pertinent. Ce principe n'est pas sans conséquence : les sites les plus anciens jouissent d'une véritable « *rente de situation* » dans leur secteur et renforcent leur position aux dépens des nouveaux arrivés [Muet & Salaün, 2001, p. 166 **(44)**].

Une analyse des sites concurrents déjà présents sur les moteurs de recherche et les annuaires peut donc s'avérer rentable. Une telle étude permet d'étudier leurs discours, leur présence sur les outils de recherche et sur des mots-clés spécifiques, leur politique de promotion (stratégie d'enchères, recours aux liens sponsorisés, au *paid inclusion*, etc.) [Saporta, 2005, p. 53, **(23)**]. Cette analyse passe également par l'observation approfondie des codes sources HTML.

2.1.2. Connaître les habitudes des internautes

Cette étape est centrale dans l'approche marketing. De la compréhension claire des besoins découlera la réflexion proprement stratégique [Muet & Salaün, 2001, p. 14 **(44)**]. Cette analyse donne naissance à une segmentation du public en fonction de ses besoins et comportements, lesquels diffèrent d'un individu à l'autre. Il s'avère donc nécessaire de former des catégories d'usagers. « *La segmentation correspond au découpage d'un marché en plusieurs sous-groupes homogènes, les segments, [...]. Chaque segment de marché correspond à une demande spécifique et homogène.* » [Muet & Salaün, 2001, p. 40 **(44)**]. Elle permet de dresser un portrait de l'utilisateur type dans ce profil et de répondre précisément à des attentes ciblées

2.2. Choix de mots-clés pertinents en fonction de la cible

Il est essentiel d'analyser les mots-clés les plus souvent choisis par les internautes pour effectuer leur recherche [Saporta, 2005, p. 4, **(23)**]. En effet, la définition de la cible du site conditionnera la politique de référencement. Ainsi, si le site s'adresse à une population particulière (professionnel, entreprise, personne âgée, femme enceinte...), il est intéressant de la spécifier dans les expressions retenues pour le référencement afin de générer un trafic qualifié [Référéneur, **(46)**].

La cible peut être définie selon trois orientations [Muet, 1998, p. 32-33, **(47)** ; Muet & Salaün, 2001, p. 113-115 **(44)**] :

- généraliste : « *on cible l'ensemble du marché, sans véritablement prendre en compte la segmentation, avec une offre de service standard. On adopte dans ce cas une stratégie de marketing indifférencié.* » [Muet & Salaün, 2001, p. 112, **(44)**]. Le site web est destiné à l'ensemble du public potentiel auquel sont offertes des prestations identiques. Cette offre standardisée caractérise un marketing de masse. Ex : La Fnac (<http://www.fnac.com>) et Surcouf (<http://www.surcouf.com>)
- spécialisée : « *on cible un segment avec une offre de service spécifique. Il s'agit d'une stratégie de marketing concentré.* » [Muet & Salaün, 2001, p. 112, **(44)**]. Le service est spécialement adapté à un segment. Cette stratégie du « créneau » occupe une part de marché réduite mais fiable (on parle aussi de « niche »).

Ex : Magrossesse.com (<http://www.magrossesse.com>), le site des femmes enceintes ou de celles qui veulent avoir un bébé ou Phosphore (<http://www.phosphore.com>), le site des 15/25 ans.

- polyvalente : le marketing adapté « *propose la même offre, mais avec des conditions de mise à disposition et d'accès qui sont modulées selon les caractéristiques de chaque segment, ou selon l'importance qu'accorde le service à tel ou tel segment.* » [Muet & Salaün, 2001, p. 114, **(44)**]. Le marketing différencié « *considère qu'à chaque segment correspond un besoin particulier, dans le fond et dans la forme.* » [Muet & Salaün, 2001, p. 115, **(44)**].

Ex : Petit Monde (<http://www.petitmonde.com>), le portail de la famille et de l'enfance, propose des rubriques adaptées à ses trois cibles : parents, professionnels et enfants. De même, sur le site principal de Flunch (<http://www.flunch.fr>), un lien redirige les jeunes visiteurs vers une page tout spécialement conçue à leur attention (couleurs vives, jeux).

Le site devra tenir compte des segments. Ainsi les pages Web axées sur la santé et s'adressant aussi bien aux professionnels qu'aux patients, devront inclure à la fois les termes médicaux spécialisés et leurs équivalents populaires quand c'est possible [Guenther, 2004, p. 49, **(48)**].

Ex. : « infarctus du myocarde » et « attaque cardiaque ».

De même, le vocabulaire employé par les jeunes ne correspond pas forcément à celui utilisé par des internautes plus âgés.

Il faut également penser à spécifier des critères géographiques (local, régional, national, international) s'ils sont importants [Andrieu, 2001, p. 70, **(02)**].

Il faudra enfin vérifier que les mots choisis soient bien utilisés sur les moteurs sans pour autant être trop concurrentiels (voir supra, chapitre 2 : « Référencement par positionnement », p. 41).

Nombreux sont les outils d'analyse et de mesure d'audience qui permettent de jauger la qualité d'un référencement.

3. Analyses et mesures d'audience

Les bibliothèques utilisent depuis longtemps les statistiques pour évaluer et améliorer leurs services [Xue, 2004, p. 181, **(30)**]. De la même manière, il est possible d'apprécier le trafic et les performances d'un site web grâce aux outils d'analyse et de mesures d'audience. Ceux-ci délivrent de multiples informations tant sur les qualités d'un référencement que sur le trafic généré, la fréquence des téléchargements ou le parcours des internautes [Chu, 2003, p. 146, **(03)**].

Ils ont été développés pour mesurer la rentabilité d'un site, qui reste un investissement commercial avant tout. L'audience en est un des indicateurs. Elle viendra en complément du volume de vente généré dans le cas d'un site de e-commerce [Chu, 2003, p. 164, **(03)**].

L'analyse de l'audience se pratique selon deux approches : la mesure côté client (*user-centric*), centrée sur l'utilisateur et la mesure côté serveur (*site-centric*), axée sur le site [Chu, 2003, p. 148, **(03)**].

3.1. Mesures côté client (*user-centric*)

Les mesures côté client, surtout adaptées aux sites à fort trafic, s'appuient sur l'analyse d'échantillons de population. Elles apportent des données d'ordre sociologique et démographique sur les internautes et permettent de décrypter les tendances générales du trafic [Chu, 2003, p. 148, **(03)**].

3.2. Mesures côté serveur (*site-centric*)

Les études axées sur le site (*site-centric*) s'intéressent avant tout à son audience. Elles sont basées sur la comptabilisation du nombre de connexions et permettent de saisir la portée du trafic généré, grâce à trois types d'analyses :

3.2.1. Analyse des logs

Un fichier *log** enregistre en temps réel toutes les activités d'un serveur web. Il fournit deux sortes d'informations : techniques (adresse IP* du visiteur, date précise de la requête, origine de la connexion, configuration du poste de l'internaute, passage des robots) et marketing (nom et taille des fichiers visités ou téléchargés, nombre de pages vues*, sites « referrers »

d'origine, pages d'entrée et de sortie). L'analyse de ces fichiers est primordiale pour évaluer la qualité du référencement. Elle peut se faire manuellement ou automatiquement par le truchement d'un logiciel. Celui-ci fera une synthèse graphique des lignes de codes pour rendre la présentation plus conviviale [Adcom, 2003, (49)].

3.2.2. Analyse à distance

Les pages dont on veut mesurer l'audience sont marquées à l'aide d'un tag. Lorsqu'une d'entre elles est visualisée par un internaute, un serveur externe d'analyse d'audience dénombre les données relatives à cette visite [Chu, 2003, p. 150, (03)].

3.2.3. Analyse des *sniffers*

Réservé aux très gros systèmes, cette technique exige l'installation d'un dispositif particulier, le *sniffer*, qui va permettre de capturer les informations transitant sur le réseau. Toutefois, il ne comptabilise pas les pages mises en cache. L'idéal est donc de combiner les analyses de logs et les techniques de tags [Chu, 2003, p. 151, (03)].

3.3. Estimation du trafic

Le positionnement n'est pas l'unique moyen d'évaluer la qualité d'un référencement. Le trafic global dérivant directement des outils de recherche majeurs (nombre de visites, mots-clés saisis par les internautes pour trouver le site) [Andrieu, 2000, p. 397, (02)] est également un bon indicateur. L'analyse est possible grâce aux fichiers logs et aux marqueurs « online » (Xiti, Estat) qui permettent une mesure directe et une mise en forme automatique de l'information [Adcom, 2003, (50)].

3.3.1. Nombre de hits

Le hit* est une action simple sur le serveur web : téléchargement d'un fichier, d'une image ou même d'un message d'erreur. Bien que le volume de hits soit un indicateur du trafic du serveur web, il ne faut pas s'y fier pour connaître le nombre de pages vues [Quatravaux, 2004, p. 259, (32) ; Xue, 2004, p. 182, (30)].

3.3.2. Nombre de pages vues

« Une page « vue » est une page considérée comme un tout cohérent du point de vue du navigateur, quel que soit le nombre d'éléments qu'elle inclut. » [Chu, 2003, p. 152, (03)].

Cet indicateur, largement utilisé, comptabilise le nombre de pages chargées sur le poste de l'internaute en phase de consultation. Il ne prend pas en compte les pages mises en cache.

3.3.3. Nombre de visites ou nombre de sessions

« Il s'agit du nombre de fois où le site a été visité pendant un intervalle de temps spécifié. » [Chu, 2003, p. 153, (03)]. Imprécises, ces données doivent être utilisées avec prudence.

3.3.4. Nombre de visiteurs

Le nombre de visiteurs permet de comptabiliser les connexions d'un même poste (et non d'un même utilisateur) sur une période donnée. Les cookies^{*23}, l'enregistrement des adresses IP, du système d'exploitation et du type de navigateur permettent d'obtenir ce type d'information.

La combinaison du nombre de hits, des pages vues et des visites fournit une bonne vision de l'usage d'un site Web [Xue, 2004, p. 182, (30)].

La provenance des visiteurs indiquera quels sont les outils de recherche utilisés ; les mots-clés permettront d'identifier le positionnement du site tandis que le trafic donnera une indication sur la bonne adéquation entre « positionnement » et « rang du site » sur les moteurs.

3.4. Identification du profil des internautes

Les outils d'analyse permettent de connaître le pays d'origine des internautes, une information utile pour une entreprise qui cherche à internationaliser ses offres. D'autres indicateurs identifient les configurations et équipements des internautes (types et versions de leur navigateur et de leur système d'exploitation, résolutions d'écran), ce qui permet d'adapter l'ergonomie et la technologie choisie en fonction des problèmes d'incompatibilité décelés [Chu, 2003, p. 154, (03)].

²³ « Variable stockée pour une durée donnée sur la machine et qui permet de conserver des informations relatives à l'identité du client et à ses actions sur le site. Lors de la première visite, le serveur enregistre ces informations dans un fichier texte qu'il conserve sur le disque dur du visiteur. Lors des visites suivantes, le serveur recherche le cookie et le configure en fonction des informations fournies » [Chu, 2003, p. 153, (03)].

3.5. Analyse du comportement des internautes

Les indicateurs de comportement identifient les pages d'entrée et de sortie des utilisateurs sur le site, données significatives pour repérer les accès privilégiés des internautes et les pages incapables de retenir leur attention. La correspondance entre pages de sortie et pages d'entrée révèle que la clientèle du site est surtout constituée d'internautes de passage [Chu, 2003, p. 155, (03)].

Les indicateurs d'itinéraires et de popularité des pages, utiles pour les activités de marketing et de promotion, dévoilent le parcours préféré des internautes, ce qui permet de déceler les pages les plus populaires (les « zones chaudes ») et les autres (les « zones froides »). Si le site web comprend un moteur de recherche interne, ces informations peuvent être associées aux mots-clés saisis par les internautes afin de déterminer si les informations demandées correspondent aux informations réellement examinées [Chu, 2003, p. 156-157, (03)].

« Jours et heures de visites renseignent sur les niveaux d'activité du serveur web pour une période spécifiée. » (nombre de pages vues et nombre de visites journalières, hebdomadaires, etc., identification des jours et des heures les plus actifs) [Chu, 2003, p. 157, (03)].

Les ratios issus des indicateurs de comportement (nombre de visites en fonction du nombre de pages vues, taux de visites à une page, fréquence de visite, nombre de pages vues à la minute) permettent de quantifier l'activité moyenne d'un site donné, son taux d'accroche, la fréquence des visites et la vitesse de lecture des pages [Chu, 2003, p. 158-159, (03)].

3.6. Performance et notoriété d'un site web

Des statistiques techniques (consommation de la bande passante, degré de fiabilité, type d'erreurs enregistrées, etc.) sont nécessaires pour évaluer la performance d'un site et l'efficacité du travail de promotion et de référencement [Chu, 2003, p. 160-162, (03)].

On distingue trois voies d'accès à un site web : l'accès direct, l'accès par liens, l'accès par les outils de recherche.

3.6.1. Accès direct

Il peut prendre trois formes : (1) saisie directe de l'URL dans le navigateur, (2) accès à partir d'un signet ou d'un favori, (3) par lien via un e-mail. L'accès direct informe sur le nombre de visiteurs fidélisés [Chu, 2003, p. 162, **(03)**].

3.6.2. Accès par liens externes

Il comptabilise les visites générées sur un site à partir de liens sur d'autres sites (*referers*). Il renseigne ainsi sur l'efficacité des partenariats et des campagnes de communication [Beranger, 2005, **(51)** ; Chu, 2003, p. 162, **(03)**].

3.6.3. Accès par les outils de recherche

Il donne des informations sur les résultats des campagnes de référencement, l'impact de chaque outil de recherche et les mots-clés spontanément utilisés par les visiteurs pour accéder à un site. Cette dernière indication peut aider à réajuster ou compléter la stratégie de référencement et à gagner de nouveaux visiteurs [Chu, 2003, p. 163, **(03)**].

A l'aune de ces données, on apprécie mieux la nécessité d'engager des démarches de mesure et d'analyse au regard de leur importance dans la gestion d'un site Web. Outre la fréquentation du site et la mesure du retour sur investissement (gain de notoriété, gain financier...), elles permettent de déceler les erreurs de conception (ergonomique, technique, etc.) mais aussi d'engager des négociations d'affiliation, de partenariat et de publicité avec d'autres sites [Chu, 2003, p. 166, **(03)**].

4. Suivi du référencement et de l'indexation

Les produits s'inscrivent dans un cycle de vie composé de quatre phases : le lancement, le développement, la maturité et le déclin. Cette vie se nourrit d'adaptations pour correspondre en continu à la demande des consommateurs [Bernadet et al., 2001, p. 64, **(45)**]. Ce truisme du marketing s'applique également au référencement (inscription sur les moteurs de recherche, développement naturel du référencement par indexation automatique, positionnement et visibilité maximale, dégradation du positionnement). Pour prévenir et contrecarrer ce processus, une veille régulière de la position du site sur les outils de recherche et l'observation de son évolution au fil des optimisations se révèlent nécessaires, au risque de le voir évincé par ses concurrents.

4.1. Détermination *a posteriori* des mots-clés

La détermination *a posteriori* des mots-clés consiste à identifier les termes saisis par les internautes pour trouver le site. Il est conseillé de faire un bilan régulier des mots-clés et expressions ayant générés le plus de trafic. Un fort taux de visite pour une expression peut résulter d'un très bon positionnement du site, de mots-clés très souvent recherchés par les internautes voire des deux tandis qu'une expression peu attractive est symptomatique d'un mauvais positionnement du site (à améliorer) ou de mots-clés dédaignés par les internautes. [WebRankInfo]. « *Cette mauvaise place devra entraîner une révision des mots-clés et une réindexation sur les moteurs.* » [Carlier, 2000, p. 181, **(01)**].

4.2. Suivi du positionnement

De même, le positionnement sur liens sponsorisés exige une attention de tous les instants. En effet, l'instabilité du classement est inhérente au système des enchères puisqu'il suffit à un éditeur de renchérir pour se positionner devant un concurrent [Saporta, 2005, p. 79, **(23)**].

4.3. Rapports de positionnement

Pour percevoir finement l'évolution du classement sur chaque mot-clé, les rapports de positionnement valent d'être confrontés. Cette comparaison peut se faire manuellement - ce qui se révèle vite fastidieux -, ou automatiquement grâce à un logiciel spécialisé qui calcule la position du site pour toutes les combinaisons de mots-clés et d'outils de recherche définis [WebRankInfo]. Une étude sur le référencement a démontré que 40 % des sondés utilisent un logiciel ou un outil en ligne pour surveiller le positionnement quand ils ne font pas appel à un référenceur [Andrieu, 2005, p. 28, **(43)**].

En ce qui concerne la mesure de l'efficacité du référencement, les professionnels se contentent de moins en moins du strict positionnement de leur page mais adoptent des préoccupations stratégiques : trafic ciblé généré par les moteurs, retour sur investissement ou analyse du chiffre d'affaires engendré sont désormais plus souvent intégrés dans la démarche de suivi [Andrieu, 2005, p. 27, **(43)**].

5. Le référencement : une méthode parmi d'autres dans le processus de promotion d'un site web

Si le référencement est une démarche essentielle pour faire connaître un site web, il est nécessaire de développer une véritable stratégie de communication pour le promouvoir efficacement et alerter de ses évolutions sur le long terme.

Plusieurs techniques peuvent être mises en place [Carlier, 2000, p. 175-176, **(01)** ; Andrieu, 2000, **(02)** ; Chu, 2003, p. 127-136, **(03)**] :

5.1. Campagnes *on-line*

5.1.1. Affiliation

Apparue aux Etats-Unis avec les sites Cdnnow.com et Amazon.com, cette technique découle du système des revendeurs de la distribution classique. Ce partenariat entre deux sites web lie deux entités : l'affilieur (celui qui propose un service en ligne) et l'affilié (celui qui va distribuer le service). Elle peut prendre deux formes : l'affiliation marchande, largement utilisée par les sites de e-commerce et la syndication* de contenu (diffusion d'une partie du contenu d'un site sur d'autres sites web) [Chu, 2003, p. 127-128, **(03)**]. L'intérêt de l'affiliation est double :

- Pour l'affilié : augmenter les revenus générés par son site ; proposer un service supplémentaire d'achat en ligne à ses visiteurs ; s'associer à la marque et bénéficier de sa notoriété et de son image.
- Pour l'affilieur : bénéficier de la notoriété de l'affilié ; augmenter la fréquentation de son site ; entretenir la visibilité de sa marque ; accroître ses commandes en ligne.

5.1.2. Cybermarketing

Ce marketing utilise « *les nouvelles technologies de l'information et de la communication comme moyen de promotion, de vente ou d'étude de marché (sites, e-mailing, mailing list, newsgroups, liens, études en ligne...).* » [Chu, 2003, p. 131, **(03)**].

Parmi les formes de marketing les plus usitées sur Internet :

- le cybermarketing direct regroupe un ensemble de techniques mises au point par un site web pour toucher des visiteurs potentiels sous la forme de campagnes d'*e-mailing*, de *newsletter* ou de *push** [Chu, 2003, p. 132, (03)] ;
- le cybermarketing viral équivaut au bouche-à-oreille dans la vie courante [Chu, 2003, p. 134, (03) ; Andrieu, 2000, (02)]
- le cybermarketing tribal rassemble des personnes aux intérêts concordants pour former une communauté virtuelle [Chu, 2003, p. 136, (03)].

5.1.3. Fils RSS

Les fils RSS* (*Really Simple Syndication* ou *Rich Site Summary* selon les sources) sont des outils supplémentaires de diffusion et de marketing. Apparus en 1999 pour la syndication de contenus entre sites, leur utilisation s'est généralisée pour diffuser des contenus en temps réel aux internautes. Ce terme « *désigne une convention de structuration permettant aux sites de diffuser des actualités au format XML.* » [Andrieu, 2004, p. 13, (52)]. Le succès des fils RSS s'explique de plusieurs façons : l'information est transmise à l'utilisateur en mode *push*, les mises à jour sont automatiques et régulières et il est possible d'y intégrer des liens sponsorisés. De plus, ils permettent de générer du trafic.

5.1.4. Publicité en ligne

Elle touche uniquement des internautes et permet un ciblage précis du public. Elle peut prendre plusieurs formes : bandeaux, bannières*, pop-ups et objets volants, page interstitielle, parrainage en ligne (*sponsoring*).

5.1.5. Webring ou chaîne de sites

Un Webring regroupe et fédère autour d'un même sujet plusieurs sites gérés par des personnes ou des entités distinctes. « *C'est une sorte de grand réseau de liens réciproques. Au lieu de créer un lien réciproque avec un seul site, vous le faites avec un groupe de sites portant sur le même thème.* » [Crowder & Crowder, 2003, p. 195, (53)]. On utilise également le terme de « chaîne » ou « anneau » pour le désigner car l'ensemble des sites forme un cercle qui peut être exploré.

« Web Art »²⁴, l'anneau francophone de l'art, regroupe 613 sites et l'«Anneau français du shareware et du freeware »²⁵ un peu plus de 170. Le site <http://www.webring.com> centralise une multitude de webrings internationaux via une classification hiérarchique inspirée des annuaires. Les structures à caractère commercial se font pour l'instant discrètes sur les Webrings.

5.1.6. Autres moyens de promotion en ligne

D'autres actions participent à la promotion d'un site : organisation de concours, participation à des groupes de discussion sur les listes, les forums* ou les groupes de *news*.

Les mini-sites web et les jeux en ligne, entièrement tournés sur la promotion d'un autre site, se révèlent pratiques pour des opérations événementielles limitées dans le temps. Enfin, le *netlinking* s'appuie sur l'échange croisé de liens entre deux sites.

5.2. Campagnes *off-line*

La promotion interne à l'entreprise, c'est-à-dire auprès des employés, peut prendre la forme d'affichages internes et de notes de services.

Les méthodes de communication traditionnelles (papiers en tête, cartes de visites, bons de livraison, factures, envoi de prospectus) participent à la diffusion de l'adresse Internet.

La promotion directe du site s'appuie également sur des campagnes d'affichage, des spots publicitaires et des communiqués de presse. La promotion indirecte passe par la participation à des salons et forums, la rédaction d'articles et des interventions dans les conférences et séminaires.

La promotion est une tâche souvent contraignante mais qui tient un rôle clé dans le succès ou l'échec d'un service web [Le Guelvouit, 1999, p. 131, **(33)**]. Une campagne purement Internet (achat de bandeaux et de liens) semble néanmoins plus efficace qu'une campagne hors média (dans la presse, à la télé ou par affichage) [Carlier, 2000, p. 179, **(01)**].

²⁴ http://www.vivance.ch/web_art/ [consulté le 26 octobre 2005]

²⁵ <http://www.sharedif.com/sharering/> [consulté le 26 octobre 2005]

CHAPITRE 4

Etude comparative : sites web de cyberlibrairies et maisons d'édition

1. Recueil des informations

Cette partie est consacrée à l'étude de sept sites web de cyberlibrairies et maisons d'édition francophones. Elle fut menée du 22 août au 16 septembre 2005.

Après avoir brièvement présenté les sites web étudiés, nous tenterons d'expliquer leur positionnement respectif – qu'il soit bon ou mauvais – sur les moteurs. Dans cet objectif, nous avons réalisé une étude fine des critères de référencement et de positionnement (utilisation des métadonnées, techniques d'optimisation et visibilité des pages). Enfin, nous aborderons la stratégie publicitaire mise en place par chacun des sites.

Des utilitaires disponibles gratuitement sur Internet ainsi que l'étude attentive des pages et de leur code HTML – affiché grâce à la fonction « source » du navigateur Internet -, permettent de cerner les méthodes d'optimisation utilisées.

Les captures d'écran liées à chaque page produit, ainsi que, selon les cas, l'en-tête HTML ou le code source complet de la page sont fournis en annexes (p. 140-160). Le tableau ci-dessous récapitule les annexes disponibles pour les pages étudiées.

Tableau 3 : Liste des annexes fournies pour chaque site étudié

	Alapage	Amazon	Chapitre	Fnac	Dunod	Eyrolles	Lavoisier
Capture d'écran	Annexe 1 p. 140	Annexe 3 p. 143	Annexe 5 p. 148	Annexe 7 p. 151	Annexe 9 p. 154	Annexe 11 p. 156	Annexe 13 p. 159
Code HTML	Annexe 2 p. 141	Annexe 4 p. 146	Annexe 6 p. 149	Annexe 8 p. 152	Annexe 10 p. 155	Annexe 12 p. 158	Annexe 14 p. 160

2. Présentation des sociétés et sites étudiés

Les sites analysés dans le cadre de cette étude sont présentés par type (cyberlibrairies et maisons d'édition) et, en leur sein, par ordre alphabétique.

2.1. Les cyberlibrairies

Les cyberlibrairies ne sont pas des éditeurs mais des revendeurs, des points de vente.

2.1.1. Alapage.fr

Alapage.fr, présent sur la Toile depuis le 30 mai 1996, appartient au fournisseur d'accès Wanadoo (France Télécom) depuis octobre 1999. Ce spécialiste des loisirs culturels et de la vente par correspondance propose plus de 3.500.000 articles appartenant à des « univers » aussi variés que les livres, cd, dvd, vidéo, jeux vidéo, logiciels, informatique, image et son, art de vivre ou billetterie. ➤ <http://www.alapage.com>

2.1.2. Amazon.fr

Fondée le 16 juillet 1995 par Jeff Bezos, un ingénieur informatique sorti de l'université de Princeton, l'entreprise de commerce électronique américaine Amazon.com, est aujourd'hui le premier libraire virtuel au monde. Avec 41 millions de clients actifs et un chiffre d'affaires de 5,8 milliards d'euros en 2004, Amazon est le leader incontesté du e-commerce. Le site, devenu un « gigantesque supermarché en ligne » [Moulin, 2005, (54)], propose désormais 31 catégories de produits.

Rien ne filtre en revanche sur les succursales du groupe alors que leur activité représente près de 45% du chiffre d'affaires en 2004 [Moulin, 2005, (54)]. La filiale française d'Amazon a ouvert le 31 août 2000. Le site propose à ce jour huit catégories de produits culturels : Livres, Livres en anglais, Musique, DVD, Vidéo, Logiciels et CD-Rom, Jeux vidéo, Image et son/micro et photo. Elle stocke plusieurs centaines de milliers de références. Le bénéfice net du groupe a augmenté de 24 millions de dollars entre le deuxième trimestre 2004 et celui de 2005 pour atteindre 76 millions de dollars. Néanmoins, une provision pour impôts et taxes de 56 millions de dollars grève ce résultat [Karayan, 2005, (55)].

➤ <http://www.amazon.fr>

2.1.3. Chapitre.com

Dernière grande cyberlibrairie francophone indépendante, Chapitre.com, créée le 20 mai 1997 par Juan Pirlot de Corbion, propose actuellement plus de 450.000 références de livres neufs, un catalogue unique en France de 3.000.000 de livres anciens, un second catalogue de plus de 15.000.000 de livres anciens grâce à ses partenaires libraires étrangers, 200.000 revues anciennes ou épuisées et de nombreuses images (gravures et affiches).

➤ <http://www.chapitre.com>

2.1.4. Fnac.com

Filiale du groupe Pinault Printemps Redoute, La Fnac, c'est d'abord 67 magasins implantés en France et 44 points de vente à l'étranger. Elle propose six types de produits : livres, disques, micro-informatique, son, vidéo et photographie.

En 2003, la société a réalisé un chiffre d'affaires de 3,812 milliards d'euros, ce qui la place parmi les vingt premières entreprises de distribution françaises et les 500 premières entreprises européennes. 18 % du CA est représenté par les livres, derrière la micro-informatique (33 %) et les disques (23 %)²⁶. ➤ <http://www.fnac.com>

Dans le top 30 français des sites les plus visités en juillet 2005 dressé par Mediamétrie²⁷, la Fnac arrive en 16^e position, loin devant Alapage (28^e) et Amazon (30^e).

2.2. Les maisons d'édition

2.2.1. Dunod.com

En ligne depuis février 1997, le catalogue de cet éditeur comporte plus de 4.500 titres principalement destinés aux étudiants et aux professionnels. Il s'est spécialisé dans cinq grandes thématiques en s'appuyant sur des collections reconnues : Sciences et Techniques (Dunod, Ediscience, ETSF), Informatique (Dunod, Microsoft Press), Economie, gestion et management, psychologie et développement personnel (InterEditions) et médico-social.

➤ <http://www.dunod.com>

²⁶

http://www.fnac.com/Magazine/entreprise_fnac/qui_est_fnac.asp?NID=0&RNID=0&SID=c2907600%2D097b%2D523c%2D4a1b%2Df3a32dba2ace&Origin=WEBREF&OrderInSession=0&UID=08f4910da%2Da604%2D6991%2D302b%2D2c66c3a7dafb&TTL=280820052208&bl=actu [consulté le 7 septembre 2005]

²⁷ http://www.mediametrie.fr/resultats.php?resultat_id=177&rubrique=net [consulté le 7 septembre 2005]

2.2.2. Eyrolles.com

Le Groupe Eyrolles est à la fois éditeur, diffuseur et libraire. Son site marchand, Eyrolles.com positionné sur la vente de livres aux professionnels, a été lancé en juillet 1999. Il s'appuie sur l'offre et l'expertise de la Librairie Eyrolles à Paris et sur sa structure de vente à distance (Dlivres). Le catalogue de 2.500 titres comprend les ouvrages de plusieurs dizaines d'éditeurs de référence français et anglo-saxons en informatique, audiovisuel et graphisme, sciences et techniques, entreprise, droit, BTP & construction, loisirs et vie quotidienne.

➤ <http://www.eyrolles.com>

2.2.3. Lavoisier.com

Créé en 1947, Lavoisier est à la fois une librairie, une agence d'abonnement et un distributeur de base de données professionnelles. Son site marchand, avant tout dédié aux professionnels, propose un catalogue de livres et cédéroms composé d'un million de titres. Avec ses quatre marques éditoriales (Tec & Doc, Hermes Science, EM Inter, Synthèse agricole), il représente un important pôle d'édition scientifique et technique. ➤ <http://www.lavoisier.fr>

2.3. Des sites dynamiques

Pour déterminer la nature des pages, nous avons observé en détail l'URL de chaque site étudié. Cet examen attentif a permis de déterminer le caractère dynamique de tous les sites : les pages sont créées à la volée lors de l'interrogation de la base de données. On les reconnaît à la présence d'un point d'interrogation dans l'URL et à l'esperluette « & » qui sépare chaque paramètre. La construction de ces sites repose sur deux technologies : l'ASP (Active Server Pages) et le PHP (Hypertext Preprocessor).

ASP (repérable par l'extension *.asp*) et PHP (repérable par l'extension *.php*) sont des langages de script exécutés du côté serveur et non du côté client (un script écrit en javascript ou une applet java s'exécutent dans le navigateur de l'internaute connecté à un site). Le contenu des pages web ASP ou PHP variera en fonction de certains paramètres (les informations stockées dans une base de données, les préférences de l'utilisateur, etc.) à la différence d'une page web « classique » (dont l'extension est *.htm* ou *.html*) qui affichera continuellement la même information (Comment ça marche, 2005, (56) et (57)).

Ces deux technologies permettent un interfaçage simple avec de nombreux systèmes de gestion de bases de données (SGBD), parmi lesquels MySQL, Oracle, PostgreSQL, Sybase, Unix dbm, etc.

Le tableau ci-dessous récapitule la technologie choisie (ASP ou PHP) par chacun des sites. Il a été possible de la mettre en évidence grâce à l’affichage d’une fiche produit, rapatriée depuis la base de données.

Tableau 4 : Technologies mises en œuvre par les sites étudiés

Technologie	Sites	Page produit	URL affichée pour la page produit
ASP	Chapitre	Saporta S. Référencement sur le net, 2005.	http://www.chapitre.com/frame_rec.asp?sessionid=63199377219493682121169250&donnee_appel=&quic ksearch=r%E9f%E9rencement&source=all
	Fnac	Saporta S. Référencement sur le net, 2005.	http://www.fnac.com/Shelf/article.asp?PRID=1622147&OrderInSession=1&Mn=3&SID=c2907600-097b-523c-4a1b-f3a32dba2ace&TTL=080920052200&Origin=FnacAff&Ra=-1&To=0&Nu=1&UID=08f4910da-a604-6991-302b-2c66c3a7dafb&Fr=0
	Dunod	Hardy J.-M. Check-list pour réussir son site web, 2004.	http://www.dunod.com/pages/ouvrages/ficheouvrage.asp?id=47518
	Lavoisier	Saporta S. Référencement sur le net, 2005.	http://www.lavoisier.fr/fr/livres/index.asp?texte=r%E9f%E9rencement&select=motcle&exact=on&togo=&support=NULL&from=
PHP	Alapage	Saporta S. Référencement sur le net, 2005.	http://www.alapage.com/mx/?id=143521124048691&donnee_appel=ALAPAGE&tp=F&type=1&l_isbn=2708133020&devise=&fulltext=&sv=X_L
	Eyrolles	Saporta S. Référencement sur le net, 2005.	http://www.eyrolles.com/Accueil/Livre/9782708133020/livre-referencement-sur-le-net.php
Non déterminé	Amazon	Saporta S. Référencement sur le net, 2005.	http://www.amazon.fr/exec/obidos/ASIN/2708133020/qid=1126123636/sr=8-2/ref=sr_8_xs_ap_i2_xgl14/402-7424880-3607302

Source : données recueillies le 7 septembre 2005 sur chaque site

Chapitre, la Fnac, Dunod et Lavoisier ont opté pour l’ASP. Alapage et Eyrolles, qui ont choisi les technologies Linux, utilisent un serveur Apache et le script PHP [Clapaud, 2005, (58) ; Eyrolles.com].

L’URL d’Amazon ne laisse pas transparaître la nature dynamique du site. Elle est dépourvue de point d’interrogation ou d’esperluette. Il n’a pas été possible de déterminer la technique utilisée (réécriture d’URL ?) pour expliquer ce résultat. En revanche, l’on sait que la société a développé des applications propriétaires sous Unix puis HP-UX en utilisant des produits du marché tel que Netscape, Oracle ou Apache [Diz, 2000, (59), Simottel, 2000, (60)].

3. Optimisation des pages

« Pour chaque requête sur un artiste ou une oeuvre, Amazon apparaît dans les premiers liens. D'ailleurs, une étude de Samaris démontre que 31 % des internautes arrivent directement sur le site marchand, sans passer chez un concurrent, alors que ce taux tombe à 10 % pour la Fnac. » [Solovieff, 2003, (61)].

Cette constatation, dressée en 2003, est-elle toujours d'actualité ? Comment peut-on l'expliquer ?

3.1. L'omniprésence d'Amazon : un positionnement stratégique

3.1.1. Méthodologie

Nous avons choisi trois syntaxes de recherche différentes (expression sans double quotes ("), lettres accentuées, majuscule au début des noms propres), correspondant à trois ouvrages portant sur Internet et le référencement :

1. Saporta référencement sur le net
2. check-list pour réussir son site web
3. Marie Prat référencement de votre site web

Ces recherches ont été lancées dans le formulaire de recherche simple, sur le web mondial de quatre grands moteurs de recherche francophones : Google France, Yahoo ! France, AltaVista France et MSN.

Le site n'est affiché que si la fiche produit correspondant à l'ouvrage recherché est présente parmi les vingt premières réponses (ou deux premières pages) pour chaque moteur.

Nous mettons donc en garde le lecteur sur la versatilité des résultats - lesquels varient en fonction de la syntaxe de recherche - et sur la difficulté de la démonstration. En effet, comme nous l'avons vu précédemment, chaque individu pose sa requête sur un sujet en fonction de processus cognitifs qui lui sont propres.

3.1.2. Résultats

Les résultats obtenus sont synthétisés dans le tableau joint en annexe 15 (p. 161). On s'aperçoit qu'Amazon domine largement ses concurrents en terme de positionnement puisqu'il est classé en première position sur les trois requêtes et ce, dans la majorité des moteurs de recherche. Chapitre, Lavoisier et Alapage n'apparaissent jamais dans les vingt premières réponses bien qu'ils vendent les ouvrages recherchés. La Fnac se positionne dans les premières pages de résultat. Dunod arrive à se placer parmi les premières réponses sur le titre qui le concerne (il ne commercialise pas les deux autres). Eyrolles enregistre de bons résultats sur les requêtes n°1 et n°3, mais demeure invisible aux premières places pour ce qui concerne la requête n°2.

3.1.3. Comparatif des sites les mieux référencés

Nous avons utilisé l'appliquatif « Outiref.com »²⁸ pour connaître le classement des sites les mieux référencés. Cet utilitaire interroge neuf outils de recherche francophones en leur soumettant le(s) mot(s) clé(s) saisi(s). Il analyse ensuite les réponses et octroie une "note" à chaque site positionné dans les dix meilleurs résultats proposés : 10 au lien affiché en premier, 9 au second, etc. jusqu'au dixième qui reçoit la note de 1. Et le tout pour chaque outil de recherche. L'addition de tous ces résultats donne un classement général des sites web les mieux positionnés pour le(s) terme(s) demandé(s). Un site peut proposer plusieurs de ses pages dans les dix liens proposés par l'outil de recherche. Ses point sont alors additionnés. Exemple : un site occupant à lui seul les première et deuxième places donne un total de 19 points (10+9) au site.

Tableau 5 : Qui est le mieux référencé pour le mot-clé Saporta référencement sur le net ?

1 -	amazon.fr	410 points
2 -	journaldunet.com	276 points
3 -	conferencevirtuelle.com	234 points
4 -	editions-organisation.com	220 points
5 -	eyrolles.com	183 points
6 -	destrucsaweb.com	166 points
7 -	yooda.com	150 points
8 -	webrankinfo.com	134 points
9 -	guideinformatique.com	130 points
10 -	sam-mag.com	109 points

Source : Outiref.com.- Consulté le 12 septembre 2005

²⁸ <http://www.outiref.com/>

Tableau 6 : Qui est le mieux référencé pour le mot-clé check-list pour réussir son site web ?

1 -	amazon.fr	370 points
2 -	ergolab.net	346 points
3 -	dunod.com	343 points
4 -	60questions.net	280 points
5 -	redaction.be	264 points
6 -	livres.journaldunet.com	184 points
7 -	amazon.ca	152 points
8 -	sam-mag.com	121 points
9 -	educnet.education.fr	115 points
10 -	qwentes.be	109 points

Source : Outiref.com.- Consulté le 12 septembre 2005

Tableau 7 : Qui est le mieux référencé pour le mot-clé Marie Prat référencement de votre site web ?

1 -	amazon.fr	375 points
2 -	destrucsaweb.com	253 points
3 -	eyrolles.com	232 points
4 -	gwix.net	213 points
5 -	guideinformatique.com	172 points
6 -	journaldunet.com	116 points
7 -	fnac.com	108 points
8 -	priceminister.com	102 points
9 -	weborama.fr	101 points
10 -	tregor.net	101 points

Source : Outiref.com.- Consulté le 14 septembre 2005

Là encore, la domination d'Amazon est prégnante puisqu'il arrive en première position sur les trois requêtes.

L'optimisation des pages peut-elle justifier cette domination sans équivoque ?

3.2. L'optimisation des pages

Afin de tenter d'expliquer au mieux le positionnement optimal d'Amazon, l'auteur a étudié les facteurs qui affectent la position du site Web sur les moteurs.

Cette analyse sur le titre des fiches produit repose en partie sur les résultats obtenus à l'aide de l'utilitaire Outiref.com sur la requête n°1 (« Saporta référencement sur le net ») et sur l'analyse des codes source des sites examinés. La page produit de Dunod concerne un autre ouvrage dans la mesure où le site n'assure pas la revente du livre de Sandrine Saporta.

3.2.1. Contenu de la balise <TITLE>

Le titre des pages est, actuellement, le critère de pertinence le plus important pour les moteurs de recherche. Il doit contenir entre sept et dix mots. La plupart des professionnels recommandent de coder les lettres accentuées en HTML (´ = é) et d'éviter les titres du type "Bienvenue sur notre site web", pas assez représentatifs du contenu des pages. Chaque page d'un site doit en outre proposer un titre différent et caractéristique du contenu qui y est proposé. La répétition de mots séparés par une virgule est à éviter au bénéfice d'une vraie phrase. Enfin dans le code HTML, il est bon de placer le titre le plus proche possible de la balise <head>.

Tableau 8 : Titre des pages produit

Site	Intitulé du titre	Taux d'optimisation
Alapage	alapage.com - Livre: REFERENCEMENT SUR LE NET	Le titre de votre page contient 45 caractère(s), et 8 mot(s). Vous avez exploité 45 % des possibilités. ★★★★★☆☆☆☆☆
Amazon	Amazon.fr : Livres: Référencement sur le net	Le titre de votre page contient 46 caractère(s), et 8 mot(s). Vous avez exploité 46 % des possibilités. ★★★★★☆☆☆☆☆
Chapitre	Votre Recherche	Le titre de votre page contient 15 caractère(s), et 2 mot(s). Vous avez exploité 15 % des possibilités. ★☆☆☆☆☆☆☆☆☆
Fnac	Fnac.com – Livres – Référencement sur le net - S. Saporta	Le titre de votre page contient 57 caractère(s), et 12 mot(s). Vous avez exploité 56 % des possibilités. ★★★★★☆☆☆☆☆
Dunod	Dunod/01 Informatique - livre Check-list pour réussir son site web : Les soixante points clés - Jean- Marc Hardy	Le titre de votre page contient 111 caractère(s), et 17 mot(s). Vous avez exploité 100 % des possibilités. ★★★★★★★★★★★★
Eyrolles	Livre Référencement sur le net - S. Saporta - Les nouvelles méthodes gagnantes - Librairie Eyrolles	Le titre de votre page contient 99 caractère(s), et 16 mot(s). Vous avez exploité 99 % des possibilités. ★★★★★★★★★★★☆☆
Lavoisier	LAVOISIER	Le titre de votre page contient 9 caractère(s), et 1 mot(s). Vous avez exploité 9 % des possibilités. ☆☆☆☆☆☆☆☆☆☆

* Page produit examinée : « Sandrine Saporta : Référencement sur le net », à l'exception de Dunod «Jean-Marc Hardy : Check-list pour réussir son site web »

Source : outiref.com .- Consulté le 8 septembre 2005

Les titres des pages produits d'Alapage, Amazon et la Fnac ont une bonne longueur avec des valeurs comprises entre 45 et 57 caractères soit 8 et 12 mots, celui d'Eyrolles est bien optimisé mais le titre de Dunod se révèle trop long. A l'exception de Chapitre et Lavoisier, les titres sont évocateurs et décrivent bien le produit présenté sur la page.

Néanmoins les lettres accentuées des titres n'ont pas été codées en HTML malgré les conseils des professionnels du référencement. Aucun des sites étudiés ne s'est plié à cette recommandation. Alapage contourne le problème en inscrivant le titre en majuscules.

Dans la plupart des cas (Chapitre, Dunod, Eyrolles, Lavoisier), le titre est idéalement placé sous la balise <head>, un peu en retrait pour Alapage et vers la fin de l'en-tête pour la Fnac et Amazon.

3.2.2. Le recours aux métadonnées

Les balises "Meta Description" et « "Meta Keywords" ont aujourd'hui beaucoup moins d'importance que par le passé dans le cadre d'un référencement. Elles n'ont une incidence que sur les moteurs de recherche qui les prennent en compte (désormais peu nombreux). Elles sont repérables à la syntaxe suivante : <meta name="description" content="description ici"> ; <meta name="keywords" content="mots-clés ici">. En règle générale, on estime qu'une balise "Meta Description" doit comporter entre 150 et 200 caractères, espaces compris. Comme pour le titre des pages, ce doit être une vraie phrase.

Quant à la balise "Meta Keywords", elle ne doit pas comporter plus de mille caractères. On peut proposer plusieurs graphies (casse, accentuation, singulier, pluriel, masculin, féminin, etc.) et intégrer les fautes d'orthographe éventuelles pour les mots-clés : référencement, référencement, REFERENCEMENT, référencement, referencement, etc. L'ordre des termes est important : les premiers ont un poids plus fort que les derniers.

La teneur de ces balises doit être différente et refléter exclusivement le contenu particulier de chaque page Web.

○ Meta Description

Tableau 9 : Meta description – Pages produits

Site	Contenu de la balise	Analyse
Alapage	REFERENCEMENT SUR LE NET	La balise "Meta Description" de votre page contient 24 caractère(s), et 4 mot(s). Vous avez exploité 12 % des possibilités. 
Amazon	Référencement sur le net, Sandrine Saporta	La balise "Meta Description" de votre page contient 42 caractère(s), et 6 mot(s). Vous avez exploité 21 % des possibilités. 
Chapitre		La balise "Meta Description" de votre page contient 0 caractère(s), et 0 mot(s). Vous avez exploité 0 % des possibilités. 
Fnac	Commandez Référencement sur le net de S. Saporta sur Fnac.com, votre magasin Fnac 24 heures sur 24, 7 jours sur 7.	La balise "Meta Description" de votre page contient 114 caractère(s), et 22 mot(s). Vous avez exploité 56 % des possibilités. 
Dunod	livre Check-list pour réussir son site web de Jean-Marc Hardy, Les soixante points clés.	La balise "Meta Description" de votre page contient 88 caractère(s), et 14 mot(s). Vous avez exploité 44 % des possibilités. 
Eyrolles	Librairie Eyrolles, vente de livres spécialisés en informatique, livres entreprise, emploi, livres management, livres developpement personnel, livres bricolage, livres droit, livres graphisme, audiovisuel sélection de livre d immobilier et urbanisme	La balise "Meta Description" de votre page contient 249 caractère(s), et 30 mot(s). Vous avez exploité 100 % des possibilités. 
Lavoisier		La balise "Meta Description" de votre page contient 0 caractère(s), et 0 mot(s). Vous avez exploité 0 % des possibilités. 

*Page produit examinée : « Sandrine Saporta : Référencement sur le net », à l'exception de Dunod «Jean-Marc Hardy : Check-list pour réussir son site web »

Source : outiref.com .- Consulté le 8 septembre 2005

Le bilan de cette analyse est hétérogène. Si la Fnac accorde une bonne longueur à la phrase de description, celle d'Eyrolles est un peu trop longue ; Alapage, Amazon et Dunod sont susceptibles de l'améliorer. Les sites Chapitre et Lavoisier l'ont totalement passée à la trappe : le code source ne contient pas de balise meta « Description ».

Or, même si ce critère est de moins en moins pris en compte par les moteurs, il mérite d'être renseigné pour chaque page Web. Les sites qui respectent cette règle proposent ou le titre (Alapage) ou le titre et l'auteur (Amazon, Dunod), plus les modalités de paiement pour la Fnac. Eyrolles s'est contenté de répéter la description de sa page d'accueil.

○ Meta Keywords

Tableau 10 : Meta Keywords – Pages produits

Site	Contenu de la balise meta Keywords	Analyse
Alapage	REFERENCEMENT SUR LE NET	La balise "Meta Keywords" de votre page contient 24 caractère(s), et 1 mot(s). Vous avez exploité 2 % des possibilités.          
Amazon	Référencement sur le net, Livres, Sandrine Saporta, E-business, Entreprise, Gestion	La balise "Meta Keywords" de votre page contient 83 caractère(s), et 11 mot(s). Vous avez exploité 8 % des possibilités.          
Chapitre		La balise "Meta Keywords" de votre page contient 0 caractère(s), et 0 mot(s). Vous avez exploité 0 % des possibilités.          
Fnac	FNAC,Fnac, fnac, livres, livre, LIVRE, bandes dessinées, bd, BD, bouquin, bouquins,guide, best-sellers, journal, librairie, libraire, roman, écrivain, polar,éducation, finance, nature, histoire, lettres, éducation, poésies, conte,enfants, informatique, langues, cuisine, littérature, scolaire, université,concours, prix, art	La balise "Meta Keywords" de votre page contient 324 caractère(s), et 73 mot(s). Vous avez exploité 32 % des possibilités.          
Dunod	livre bureautique, livres bureautique	La balise "Meta Keywords" de votre page contient 37 caractère(s), et 3 mot(s). Vous avez exploité 3 % des possibilités.          
Eyrolles	libraire, éditeur, livres entreprise, Eyrolles, ouvrages management, emploi, marketing, ressources humaines, professionnels, livres de commerce, nouvelles technologies, stratégie, finance, droit, études, vie pratique, Livre, Référencement sur le net, S. Saporta, Les nouvelles méthodes gagnantes, marketing web	La balise "Meta Keywords" de votre page contient 331 caractère(s), et 47 mot(s). Vous avez exploité 33 % des possibilités.          
Lavoisier		La balise "Meta Keywords" de votre page contient 0 caractère(s), et 0 mot(s). Vous avez exploité 0 % des possibilités.          

* Page produit examinée : « Sandrine Saporta : Référencement sur le net », à l'exception de Dunod «Jean-Marc Hardy : Check-list pour réussir son site web » Source : outiref.com - Consulté le 8 septembre 2005

La balise meta « Keywords » ne semble guère retenir l'attention des webmestres. En effet, totalement ignorée par Chapitre et Lavoisier, elle reste sous-exploitée dans les autres pages (2 à 33 %). C'est une fois encore Eyrolles qui optimise au mieux cette balise. Au niveau du contenu : Alapage reprend le titre de l'ouvrage, Amazon le titre et l'auteur auxquels s'ajoutent quelques mots-clés. La Fnac offre une répétition de mots-clés sans rapport avec le contenu, Dunod se contente du minimum : il décline le concept au singulier/pluriel. Eyrolles propose des mots-clés généraux suivis du titre et du nom de l'auteur du livre puis d'un mot-clé spécifique à l'ouvrage (« marketing web »).

○ Meta NAME

Ce tableau récapitule les META NAME employées dans les pages produits étudiées :

Tableau 11 : Emploi des meta name dans les pages produits

Nom du site	Alapage	Amazon	Chapitre	Fnac	Dunod	Eyrolles	Lavoisier
Balises	description keywords	description keywords	-	ms-locale category description keywords robots revisit-after	description title keywords	Keywords description	-

Source : code source HTML des pages étudiées

Comme on peut le constater, seul la Fnac a fait un réel effort d'insertion de métadonnées. En sus des classiques meta name « Description » et « Keywords », la page comporte des métadonnées destinées aux robots (robots, revisit-after).

Ces deux balises doivent prendre place après la balise <title>...</title> et avant la balise de fin d'en-tête (</head>). La balise <title>, doit conserver sa première position car, comme nous l'avons vu auparavant, ce champ est plus important que les balises <meta> en tant que critère de pertinence. Pourtant les codes source des pages de la Fnac et d'Amazon ne respectent pas cette règle.

○ Meta HTTP-EQUIV

Ces balises ne sont prises en compte par aucun moteur de recherche majeur. En revanche, elles apportent des informations complémentaires au navigateur.

Tableau 12 : Emploi des meta HTTP-EQUIV dans les pages produits

Nom du site	Alapage	Amazon	Chapitre	Fnac	Dunod	Eyrolles	Lavoisier
Balises	-	-	-	PICS-Label Content-Type	Content-Type	Content-Type Content-Script-Type Content-Style-Type Content-Language	-

Source : code source HTML des pages étudiées

La meta « http-equiv="Content-Type » utilisée par la Fnac, Dunod et Eyrolles spécifie l'encodage général des caractères utilisés dans la page, dans les trois cas qui nous intéressent l'ISO 8859-1 ou ISO Latin 1.

La norme PICS-LABEL (*Platform for Internet Content Selection* ou PICS) appliquée par la Fnac permet l'utilisation de navigateurs interdisant l'accès à certaines pages dans l'objectif de protéger les mineurs sur Internet [Adcom, 2003, (62)].

Eyrolles utilise également d'autres meta :

- META http-equiv="Content-Script-Type" : nécessaire si l'on utilise des scripts dans des attributs des balises du corps de la page.
- META http-equiv="Content-Style-Type" : nécessaire si l'on utilise un ou des attributs de style directement dans les balises du corps de la page et que la valeur de ces attributs soit conforme aux "Cascading Style Sheet".
- META http-equiv="Content-Language" : spécifie la langue utilisée

En revanche, aucun des sites n'a pris la peine d'optimiser ses pages grâce à l'insertion de métadonnées Dublin Core. Cette absence dans des sites e-commerce d'importance nationale est toutefois caractéristique d'une tendance plus générale qui se vérifie sur Internet : le recours aux métadonnées Dublin Core reste peu usité.

3.2.3. Texte visible de la page

On préconise en général au moins 300 caractères (une lettre accentuée équivalant à un caractère) en texte visible pour qu'une page soit "réactive" par rapport aux moteurs de recherche [Outiref.com].

C'est pourquoi il faut éviter les pages trop graphiques qui risquent d'être mal classées sur les moteurs de recherche. Ces derniers sont par ailleurs très sensibles à la présence de mots-clés importants dans le texte visible. Les résultats de cette analyse peuvent être consultés en annexe 16 (p. 162). A l'exception de Chapitre (82 caractères) et Lavoisier (1 caractère !), le contenu visible de la page est optimisé. Les autres pages comprennent entre 1385 caractères (Dunod) et 5972 caractères (Eyrolles).

3.2.4. Contenu invisible de la page

Tableau 13 : Contenu des pages produits

Alapage	Nombre d'images : 41 Nombre de liens : 46 Présence de FRAMES : Non Présence de FLASH : Non Présence de COMMENTAIRES : Oui Présence de BALISES "ALT" : Non	Dunod	Nombre d'images : 29 Nombre de liens : 27 Présence de FRAMES : Non Présence de FLASH : Non Présence de COMMENTAIRES : Oui Présence de BALISES "ALT" : Non
Amazon	Nombre d'images : 36 Nombre de liens : 127 Présence de FRAMES : Non Présence de FLASH : Non Présence de COMMENTAIRES : Oui Présence de BALISES "ALT" : Oui	Eyrolles	Nombre d'images : 24 Nombre de liens : 152 Présence de FRAMES : Non Présence de FLASH : Non Présence de COMMENTAIRES : Oui Présence de BALISES "ALT" : Oui
Chapitre	Nombre d'images : 1 Nombre de liens : 0 Présence de FRAMES : Oui Présence de FLASH : Non Présence de COMMENTAIRES : Oui Présence de BALISES "ALT" : Non	Lavoisier	Nombre d'images : 0 Nombre de liens : 0 Présence de FRAMES : Oui Présence de FLASH : Non Présence de COMMENTAIRES : Non Présence de BALISES "ALT" : Non
Fnac	Nombre d'images : 30 Nombre de liens : 73 Présence de FRAMES : Non Présence de FLASH : Non Présence de COMMENTAIRES : Oui Présence de BALISES "ALT" : Non		

Source : Outiref.com.- Consulté le 12 septembre 2005

○ Facteurs bloquants

Si certains moteurs (comme Google) savent de mieux en mieux lire le format Flash et suivre les liens, les *frames* posent parfois de gros soucis de référencement. Il convient d'insérer un texte par défaut dans la balise <noframes>...</noframes>.

Aucune des pages étudiées n'est basée sur du Flash. En revanche, les codes source considérés, à l'exception de celui de Lavoisier, démontrent l'emploi généralisé du Javascript. En outre, les sites de Chapitre et Lavoisier sont construits avec des *frames*, ce qui grève fortement la qualité de leur référencement. Ils sont congénitalement handicapés par rapport aux autres pages.

○ Balise ALT

Les options ALT sont en règle générale prises en compte par les moteurs de recherche, même si leur importance est assez faible. Le texte qui y est inséré peut cependant servir aux moteurs de recherche d'image pour indexer et reconnaître les fichiers graphiques.

A l'exception des pages d'Amazon et d'Eyrolles, qui associent une balise ALT « Référencement sur le net » à la présentation de la couverture de l'ouvrage, ces balises ne sont pas mises à profit. Bien que présentes dans le code source des pages Dunod et Alapage, n'étant pas renseignées, elles s'avèrent inutiles.

3.2.5. Mise en exergue et valeur ajoutée

Tableau 14 : Mise en exergue du contenu

	Moyens de mise en exergue	Valeur ajoutée
Alapage	Titre en gras majuscule Lien sur nom de l'auteur	« L'avis des clients »
Amazon	Titre <H1> en gras Liens sur nom de l'auteur et titre de l'ouvrage	« Les internautes ayant acheté (titre de l'ouvrage) ont également acheté » « Les internautes intéressés par cet article peuvent être intéressés par : » « Chroniques et points de vue » « Les internautes ayant acheté des titres de (nom de l'auteur) ont également acheté des titres de » « Rechercher des livres semblables à (titre de l'ouvrage) par sujet »
Chapitre		
Fnac	Titre <H5> en gras Liens sur nom de l'auteur et titre de l'ouvrage	« Les internautes ayant acheté cet article ont également acheté : » « Vos avis ».
Dunod		
Eyrolles	Titre <H2>, gras ou Liens sur nom de l'auteur	« Extraits du livre » « Avis des lecteurs » « Les internautes qui ont acheté ce livre ont aussi acheté » « Consultez aussi »
Lavoisier	Lien sur nom de l'auteur	

Source : codes source des pages HTML .- Consultés le 16 septembre 2005

Au-delà des classiques notices d'ouvrages, résumés et sommaires, certains sites apportent une réelle valeur ajoutée à leur page produit grâce à des rubriques originales. C'est ainsi qu'Alapage, Amazon, la Fnac et Eyrolles invitent les lecteurs à laisser leurs avis. Les mêmes, à l'exception d'Alapage orientent les internautes dans leurs achats ou stimulent leurs centres d'intérêt (Amazon, Eyrolles). Eyrolles est le seul à proposer des extraits de l'ouvrage.

3.2.6. Contenu de l'URL

Tableau 15 : URL des pages produits

Alapage	http://www.alapage.com/mx/?id=143521124048691&donnee_appel=ALAPAGE&tp=F&type=1&l_isbn=2708133020&devise=&fulltext=&sv=X_L	L'URL fait 121 caractères.
Amazon	http://www.amazon.fr/exec/obidos/ASIN/2708133020/qid=1126123636/sr=8-2/ref=sr_8_xs_ap_i2_xgl14/402-7424880-3607302	L'URL fait 114 caractères.
Chapitre	http://www.chapitre.com/frame_rec.asp?sessionid=63199377219493682121169250&donnee_appel=&quicksearch=r%E9f%E9rence ment&source=all	L'URL fait 129 caractères.
Fnac	http://www.fnac.com/Shelf/article.asp?PRID=1622147&OrderInSession=1&Mn=3&SID=c2907600-097b-523c-4a1b-f3a32dba2ace&TTL=080920052200&Origin=FnacAff&Ra=-1&To=0&Nu=1&UID=08f4910da-a604-6991-302b-2c66c3a7dafb&Fr=0	L'URL fait 208 caractères.
Dunod	http://www.dunod.com/pages/ouvrages/ficheouvrage.asp?id=47518	L'URL fait 61 caractères.
Eyrolles	http://www.eyrolles.com/Accueil/Livre/9782708133020/livre-referencement-sur-le-net.php	L'URL fait 86 caractères.
Lavoisier	http://www.lavoisier.fr/fr/livres/index.asp?texte=r%E9f%E9rence ment&select=motcle&exact=on&togo=&support=NULL&from=	L'URL fait 115 caractères.

Source : outiref.com.- Consulté le 8 septembre 2005

Nous l'avons vu, les pages dynamiques sont plus difficiles à indexer que les autres en raison de la syntaxe même de leur URL. Pourtant, seul Eyrolles semble avoir totalement optimisé l'adresse de sa page produit. Celle-ci est très évocatrice puisqu'elle reprend le titre de l'ouvrage. Celle d'Amazon, en revanche, laisse perplexe. Bien que les signes typiques d'une page dynamique ne soient pas évidents, l'URL demeure pour le moins sibylline (succession de chiffres). Les pages produits de tous les autres sites comportent des points d'interrogation et/ou esperluette, rendant leur URL « exotique » au regard des moteurs de recherche. Chapitre affiche le contenu de sa base de données dans un cadre* (*frame*), l'URL ne changeant pas, quelque soit la fiche produit consultée.

Mis à part l'URL de Dunod, toutes les adresses dépassent allégrement les 70 caractères préconisés.

3.2.7. Indice de densité des mots-clés

Plus la densité d'un mot-clé est importante sur une page Web, plus ce dernier sera réactif. La densité de mots-clés correspond au nombre d'occurrences d'un mot sur le nombre total de mots dans la page.

Les résultats, synthétisés dans l'annexe 17 (p. 163), montrent que la densité du mot-clé « référencement » est compris entre 1,69 % (Alapage) et 2,77 % (Fnac), en fonction des pages. La Fnac et Dunod (3,01 % sur le mot-clé « site web » et 2,42 % sur celui de « réussir site web ») se révèlent les mieux optimisés, les bonnes limites pour l'indice de densité se situant entre 2,5 et 5,5 % [Andrieu, 2005, p. 4, **(21)**]. Chapitre et Lavoisier jouent toujours les mauvais élèves mais cette situation s'explique en grande partie par l'emploi des *frames*.

4. Visibilité et positionnement des sites sur les outils de recherche

4.1. Visibilité des pages

Visiref calcule l'indice de visibilité des sites. Cet applicatif développé par Brioude Internet et Perl-Pro va interroger les principaux moteurs de recherche francophones (Google France, Yahoo France, Msn France, Aol France, Voila Moteur) en leur soumettant le(s) mot(s) clé(s) saisi(s). Il analyse les réponses et octroie une "note" à chaque site positionné dans les dix meilleurs résultats proposés : 10 au lien affiché en premier, 9 au second, etc. jusqu'au dixième qui reçoit la note de 1. Un système de "bonus" est utilisé pour les trois premières positions. Les résultats fournis sont également pondérés en tenant compte des parts de trafic de chaque outil de recherche. Plus le positionnement sur un outil de recherche "important" sera bon, meilleur sera le classement sur Visiref. L'addition de tous ces résultats donne un classement général des sites web les mieux positionnés pour le(s) terme(s) demandé(s).

Tableau 16 : Classement de visibilité des pages produits en fonction des requêtes

Requête	Site	Classement Visiref	Visipoints
N°1	Alapage	>64	
	Amazon	1	2220
	Chapitre	>64	
	Fnac	21	174
	Eyrolles	2	2072
	Lavoisier	>64	
N°2	Alapage	>66	
	Amazon	1	2220
	Chapitre	>68	
	Fnac	>73	
	Dunod	2	2072
	Eyrolles	23	148
	Lavoisier	>74	
N°3	Alapage	24	132
	Amazon	1	2220
	Chapitre	>65	
	Fnac	21	162
	Eyrolles	3	1998
	Lavoisier	>65	

Source : Visiref.com .- Consulté le 14 septembre 2005

Quelle que soit la requête, Amazon arrive systématiquement en tête de classement, ce qui corrobore les données précédentes. Eyrolles et Dunod (pour l'ouvrage qui le concerne) sont également bien positionnés. Les autres sites seraient bien inspirés d'améliorer leur visibilité pour générer davantage de trafic via les moteurs.

4.2. Nombre de pages indexées

Pour connaître le nombre de pages indexées pour un site en particulier, il suffit de restreindre la recherche à un domaine. Pour cela il faut utiliser la syntaxe « site : domaine.com » sur la plupart des moteurs. Ces derniers renvoient alors les pages Web appartenant au site spécifié.

Tableau 17 : Nombre de pages indexées

	Google	Yahoo !	MSN	Total	Classement
Alapage	823 000	953 000	188	1 776 188	4
Amazon	3 250 000	20 800 000	141 338	24 191 338	1
Chapitre	414 000	771 000	5 552	1 190 552	5
Fnac	3 750 000	6 840 000	14 809	10 604 809	3
Dunod	84 700	7 510	7 434	99 644	7
Eyrolles	4 090 000	10 100 000	2 885	14 192 885	2
Lavoisier	309 000	30 100	1 608	340 708	6

Source : moteurs de recherche.- Consultés le 9 septembre 2005

On note une forte disparité dans le nombre de pages indexées en fonction des moteurs. Toutefois, on peut remarquer que, au total, Amazon arrive une fois de plus en tête, suivi d'Eyrolles puis de la Fnac. Les autres sites sont loin derrière, Dunod fermant le peloton.

4.3. PageRank© de Google

Le PageRank© (PR) des sites Web étudiés a été mesuré grâce à la barre d'outils de Google. Deux autres utilitaires ont été employés afin de vérifier les premiers résultats : <http://www.mygooglepagerank.com> et <http://www.yagoort.org/tools>.

Tableau 18 : PageRank© Google des sites étudiés

Sites Web	Alapage	Amazon	Chapitre	Fnac	Dunod	Eyrolles	Lavoisier
PR Google	7/10	0/10	7/10	7/10	3/10	7/10	7/10
PR mygoogle	7/10	8/10	7/10	7/10	6/10	7/10	7/10
PR Yagoort	7/10	7-8/10	7/10	7/10	6/10	7/10	6-7/10

Sources : <http://www.mygooglepagerank.com>, <http://www.yagoort.org/tools> et Google Tool Bar.- Consultés le 27 août 2005

Les trois outils utilisés pour mesurer le PageRank© de Google fournissent dans l'ensemble des données identiques, le PR de la plupart des sites se situant entre 6 : popularité moyenne (Dunod) et 8 : bonne popularité (Amazon). On note en revanche deux résultats étranges : le PR de 0/10 d'Amazon et le PR de 3/10 (popularité très mauvaise) de Dunod, tous deux fournis par la barre d'outil de Google. Amazon a-t-il passé des accords avec Google pour masquer son PR ? S'agit-il d'un bug ? De même, le faible PR de Dunod entre en contradiction avec les résultats délivrés par les deux autres applicatifs qui le gratifient d'un PR de 6/10. Dans l'état actuel de nos connaissances, il est difficile d'apporter une explication à ce phénomène.

Notons toutefois que le PR de Google ne reflète pas la popularité d'un site Web. D'autres composants entrent dans le calcul. Un faible PR n'implique pas nécessairement une faible popularité.

4.4. Indice de popularité

L'indice de popularité (IP) d'un site se mesure au nombre de pages web pointant vers lui. Il est calculé à la fois sur un mode quantitatif (nombre de liens) et qualitatif (qualité de ces liens). Il est pris en considération par de nombreux moteurs dans leurs classements de pertinence.

Tableau 19 : Indice de popularité des sites étudiés

	Alapage	Amazon	Chapitre	Fnac	Dunod	Eyrolles	Lavoisier
Google	300	44 600	0	21 100	168	163	1 040
Yahoo	7 440	6 120 000	6 700	2 000 000	7 400	445 000	17 000
MSN	1 634	23 934	123	34 358	3 166	4 892	9 083

Source : Outiref.com. – Consulté le 22 août 2005

Amazon et la Fnac sont les deux sites les plus populaires de notre panel. La position de leader d'Amazon est renforcée par sa notoriété sur Google et Yahoo, qui nous l'avons vu, drainent la grande majorité du trafic Internet.

La même démarche appliquée aux fiches produits se révèle décevante et inutile dans la mesure où ces pages, dynamiques et susceptibles de disparaître ou changer à tout moment, font rarement l'objet de liens directs.

Les résultats de cette étude suggèrent qu'un bon classement dans la page de résultats d'un moteur de recherche dépend d'une combinaison de facteurs. Les sites avec le meilleur classement ont souvent les meilleures valeurs sur les critères testés. Les sept facteurs considérés dans cette étude – contenu de la balise <TITLE>, recours aux métadonnées, densité de mots-clés et URL, visibilité des pages, nombre de pages indexées, PR de Google, popularité des sites Web – ont une grande influence.

Cependant, avec le secret qui entoure l'algorithme de classement et le développement rapide des moteurs de recherche, il est difficile de déterminer avec certitude et précision ce qui amène un site web au sommet des pages de résultats.

5. Le plan marketing des sites étudiés

5.1. Liens sponsorisés (voir annexe 18, p. 164)

Moklic²⁹ est un utilitaire mis gratuitement à la disposition des internautes. Il laisse entrevoir la stratégie marketing appliquée par les cyberlibrairies et maisons d'édition en ligne.

Un utilitaire proposé sur Miva.com permet de suivre les enchères. De même, sur Overture, un lien (relativement discret) placé en haut à droite de la page de recherche permet de connaître les offres maximales en cours. Par contre, nous n'avons pas trouvé les prix pratiqués aux enchères du programme AdWords de Google.

Nous avons choisi huit expressions au hasard liées à l'activité et aux produits des cyberlibrairies et maisons d'édition dans l'optique de cerner leur politique d'achat de mots-clés.

La vision que nous proposons ici est on ne peut plus éphémère et volatile. Les enchères et les mots-clés évoluant chaque jour, il faut considérer ce tableau comme une image arrêtée de l'état du marché des liens sponsorisés sur des mots-clés précis à un instant T. Il peut néanmoins nous permettre de dégager quelques généralités.

Les cyberlibrairies s'octroient une bonne visibilité grâce à l'achat de mots-clés aux enchères. Toutefois leurs stratégies divergent. Si Amazon et la Fnac se contentent d'être présents sur Google, Chapitre s'impose sur l'ensemble des régies tandis qu'Alapage se positionne essentiellement sur les réseaux Miva et Overture.

Chapitre et Alapage déploient une stratégie de promotion relativement agressive (sur les huit expressions testées et l'ensemble des régies, Chapitre comptabilise 22 liens sponsorisés contre 11 pour Alapage). Ils sont prêts à payer cher un clic sur des mots-clés très généraux comme « livre » ou « livres », susceptibles de générer un trafic important.

²⁹ <http://www.moklic.com>

5.2. Affiliation

Ce concept créé par Amazon consiste à rétribuer des sites supports de bannières publicitaires en fonction des ventes qu'ils génèrent. Tous les types de sites peuvent participer : médias, moteurs de recherche, portails, sites marchands ou encore sites de particuliers ou d'associations. Tous les sites examinés ont développé de tels programmes.

Sur son site, Alapage admet que sa stratégie globale de communication repose sur sa politique d'affiliation. Le cyberlibraire compte à ce jour plus de 4.000 partenaires issus de secteurs d'activité hétéroclites tels que Radio France, le groupe NRJ, La Recherche, Le Magazine Littéraire, Voila, Orange, Le Routard... Toute vente générée est reversée au partenaire sous la forme d'une commission si le montant cumulé des ventes est supérieur à 80 euros.

Xavier Garambois, dirigeant d'Amazon France reconnaît que le « *premier poste de dépenses correspond à notre programme d'affiliation. Ensuite vient l'achat de mots-clés.* » [Karayan, 2005, (55)]. Entre 500.000 et 600.000 sites web ont souscrit au programme Partenaires d'Amazon. Ce dernier reverse une commission allant de 2,5 % à 7,5 %, voire 11 % (pour le modèle de paiement en paliers) sur chaque vente générée. Les partenaires sont identifiables à une petite icône ou à la phrase "En association avec amazon.fr".

Chapitre reverse 5 € par vente réalisée dans une période de 30 jours après la visite du site. L'inscription s'effectue sur la plate-forme française Cibleclick.com.

Avec le programme de la Fnac, les affiliés peuvent engranger jusqu'à 8 % de commission. Le seuil minimum de facturation est néanmoins fixé à 50 € HT.

Depuis 2000, Eyrolles a mis en place de nombreux partenariats. On retrouve ainsi son offre de librairie sur des sites d'information ou de services. Une commission est reversée chaque mois à l'affilié si les ventes générées dépassent les 100 € HT. Dans le cas contraire, ce montant est reporté au trimestre suivant. Parmi ses plus prestigieux partenaires, on peut citer Build2pro, Cadremploi, CadresOnline, Caloga, Cyberarchi, PharmaNetwork, etc.

Lavoisier et Dunod restent discrets sur leur politique de partenariat et renvoient les potentiels affiliés vers les services appropriés pour plus d'information.

A l'image d'Alapage ou Amazon, l'affilieur met très souvent à la disposition de ses partenaires des outils propriétaires (moteur de recherche, visuels génériques ou d'actualité tels que vignettes, bannières et bandeaux) facilitant l'animation des rubriques d'achat.

Tous les affiliés participent à augmenter la popularité des sites dont ils sont les partenaires. Ils jouent donc un rôle non négligeable dans le positionnement de ces sites en impactant leur indice de popularité.

5.3. Des sites supports de publicité

Dans le top 50 des sites supports de publicité de juillet 2005³⁰, Alapage se classe en 18^e position, le rang des autres sites dépassant la barre des 50, ils n'apparaissent pas dans l'enquête de Mediamétrie.

5.4. Fils RSS

Si le site américain d'Amazon propose la liste de ses derniers produits³¹, Eyrolles est le seul site de l'étude à offrir - depuis mars 2005 - des fils d'information RSS à ses partenaires. La syndication permet à ces derniers de surveiller les nouveautés du catalogue et de mettre à jour et animer la rubrique "Livres" (ou librairie, bibliothèque, etc.) de leurs sites.

³⁰ http://www.mediametrie.fr/resultats.php?resultat_id=177&rubrique=net [consulté le 28 août 2005]

³¹ <http://www.amazon.com/exec/obidos/subst/xs/syndicate.html/102-7515222-8848904> [consulté le 15 septembre 2005]

Au terme de cette étude, nous pouvons justifier le positionnement « naturel » et optimal d'Amazon. Il repose sur une combinaison de facteurs :

- des critères d'optimisation maximaux pour le référencement :
 - Bien que mal placé dans l'en-tête, chaque page comporte le titre (<TITLE>) exact de l'ouvrage.
 - Les balises méta nécessaires (Description et Keywords) sont renseignées et comportent toutes deux le titre et le nom de l'auteur de l'ouvrage auxquels correspond la fiche produit.
 - Le titre HTML est mis en exergue (balise H et gras), ce qui est un atout majeur dans les critères de pertinence.
 - Le contenu (balises ALT, liens internes) est optimisé.
 - Aucun facteur bloquant (ni cadres, ni Flash) ne complique l'indexation.
 - L'indice de densité des mots-clés, s'il n'est pas parfait, demeure néanmoins correct.
- Le site est très populaire pour Google et Yahoo, les deux principaux moteurs de recherche francophones.
- L'ancienneté : le site est présent de longue date dans les moteurs. Or, plus le site est ancien dans l'index d'un moteur, plus il a de chance de sortir dans les premières pages de résultat.
- Le nombre de pages indexées témoigne d'une bonne visibilité : 3.250.000 pages indexées à ce jour dans Google et 20.800.000 dans Yahoo.
- Les avis des lecteurs et autres informations à forte valeur ajoutée renforcent le contenu déjà existant.
- Son positionnement en tête du classement sur les pages de résultats lui permet de capter les internautes.

- La stratégie marketing basée sur un programme d'affiliation dynamique et l'achat de mots-clés ciblés sur Google participe à accroître sa popularité et sa visibilité.

Dans l'ensemble, les pages des concurrents d'Amazon sont plutôt bien conçues. Alapage et la Fnac ont tous deux un titre évocateur, les méta nécessaires, un contenu relativement bien optimisé (malgré l'absence d'attribut ALT), un niveau d'indexation correct et pour la Fnac une forte popularité sur Google. Outre son titre, trop long, la page produit de Dunod peut être améliorée. Le site gagnerait surtout à accroître sa popularité, insuffisante, via l'échange de liens.

Eyrolles est le seul à avoir « travaillé » l'URL de ses pages produits et cela paye puisqu'il se place en deuxième position derrière Amazon en terme de nombre de pages indexées. Ses pages sont un exemple d'optimisation réussie.

A contrario, Lavoisier et Chapitre, construits avec des cadres, patissent d'une conception hasardeuse à la base. Ils ne sont pas du tout optimisés (pas de description, pas de métadonnées, pas de contenu réactif, mauvaise visibilité), ce qui joue en leur défaveur lors de l'indexation des pages. Chapitre se voit donc contraint de développer une démarche de promotion payante très active pour assurer sa visibilité.

Enfin, d'autres facteurs, moins évidents au premier abord, expliquent également l'hégémonie d'Amazon :

La société (Amazon.com) jouit d'une notoriété internationale. Le mot-clé « Amazon » arrive en 101^e position dans le top 500 des mots les plus recherchés sur le moteur Searchwords³² (données du 21 août 2005). C'est donc une enseigne reconnue et populaire, qui ne l'oublions pas, propose une offre de produits vaste et diversifiée. Et dans ce dernier paramètre se cache peut-être la raison fondamentale de sa popularité et de son omniprésence. Les maisons d'édition peuvent difficilement concurrencer les entreprises de vente de produits pluriculturels. Leur offre est, en effet, beaucoup plus réduite.

³² <http://www.searchwords.com> [consulté le 21 août 2005]

CONCLUSION

Le positionnement d'un site sur les pages de résultats des moteurs de recherche est le garant de sa visibilité. Il dépend essentiellement de la nature et de la qualité du référencement. Cette dernière opération conditionne le volume et la qualification du trafic drainé sur un site.

Par-delà les nuances de vocabulaire, la question du référencement voit se confronter deux visions de l'accès à l'information : la vision documentaire classique, chère aux documentalistes, et la vision marketing, préférée par les sites de e-commerce. Cette dernière semble s'imposer dans l'océan d'information tentaculaire qu'est devenu Internet.

Dans un environnement de plus en plus concurrentiel, être présent sur les moteurs de recherche ne suffit plus, les sites de e-commerce en particulier, doivent impérativement se classer dans les premières places pour assurer leur survie économique. Pour y parvenir, l'optimisation des pages n'est pas toujours suffisante. Les nouvelles offres payantes, dont on peut supputer avec raison l'accroissement exponentiel, permettent de contourner cette difficulté. Un bon référencement n'est que provisoire, les résultats obtenus sont fugaces et relatifs. En raison de l'évolution rapide et anarchique d'Internet et de la création de milliers de sites chaque jour, sa gestion et son suivi dans la durée se révèlent indispensables.

Cet exposé a démontré que les réponses affichées à la suite d'une requête dépendent étroitement des critères de pertinence et de classement des moteurs de recherche. C'est pourquoi, pour effectuer une recherche fructueuse et discriminer au mieux les informations, il appartient aux professionnels de l'information et de la documentation de bien connaître les moteurs de recherche, de se tenir informés de l'évolution des nouvelles technologies et des techniques de référencement, toujours plus perfectionnées (la syndication n'en est qu'à ses balbutiements) en gardant toutefois à l'esprit que la visibilité d'un site ne préjuge pas nécessairement de sa pertinence mais témoigne aussi d'une maîtrise plus affirmée des méthodes de *search marketing**.

BIBLIOGRAPHIE

Présentation de la bibliographie

La bibliographie comprend 62 références.

La collecte de documents a cessé en septembre 2005.

Cette bibliographie ne se prétend nullement exhaustive. Elle présente une sélection des sources qui reflètent les interrogations et les besoins documentaires rencontrés pour l'élaboration de ce mémoire.

Deux présentations sont proposées dans le souci de faciliter une recherche bibliographique : par thèmes et par noms d'auteurs.

○ Bibliographie thématique/analytique commentée

Numérotée et classée par ordre d'apparition des renvois dans le texte, la bibliographie analytique est présentée en premier. Le plan de classement des notices suit le plan du mémoire et celles-ci sont accompagnées d'un résumé. Un numéro entre parenthèse « **(01)** » est attribué à chaque notice. Ces numéros servent de référence dans le corps du texte du mémoire.

Les notices sont suivies d'un résumé provenant de quatre origines différentes. Ce sont soit :

- les résumés d'auteur des publications,
- les résumés extraits du Bulletin Bibliographique de l'INTD (Institut National des Techniques de la Documentation). Ils ont été conservés en l'état dans la majorité des cas,
- des résumés que nous avons nous-même rédigés pour l'occasion et notés *A.M.*

○ Bibliographie alphabétique des noms d'auteurs

La bibliographie par ordre alphabétique des noms d'auteurs (personnes physiques ou morales), et pour le même auteur, par ordre chronologique inverse (du plus récent au plus ancien) permettra un repérage facile et rapide d'une référence bibliographique. Pour une même année ou un même mois, c'est l'ordre alphabétique du titre qui prévaut. Les notices ne sont pas accompagnées de résumés.

○ Bibliographie complémentaire

Enfin, les références n'ayant pas servi directement lors de la rédaction du mémoire sont néanmoins présentées au titre de sources pertinentes sur le sujet dans une bibliographie complémentaire.

Afin de faciliter l'identification des documents, quelques icônes ont été utilisées :



= monographie



= article



= ressource électronique



= mémoire INTD

Bibliographie analytique

CHAPITRE 1 – QU’EST-CE QU’UN MOTEUR DE RECHERCHE ?

○ Présentation des outils de recherche

(01) CARLIER, Richard. *Profession webmaster*. Paris : « Les Echos » Ed., 2000, 255 p. 

Le webmaster est chargé de gérer un site Internet. L’auteur, qui fait partie de la profession, explore des thèmes aussi divers que la conception graphique des sites, l’exploitation des bases de données, la sécurité, le cybermarketing, la veille technologique, la formation ou encore le recrutement. Des références permettent d’approfondir chaque point. Le chapitre 14 (p. 169-182) est entièrement consacré au référencement. A.M.

(02) ANDRIEU, Olivier. *Créer du trafic sur son site web : conception, référencement, promotion*. Paris : Eyrolles, 2000, 532 p. 

Le référencement des sites web sur les outils de recherche majeurs ne s’improvise pas. Cet ouvrage, une référence en la matière, expose de manière méthodique les techniques à mettre en oeuvre lors de la conception, du référencement (Chapitre 4, p. 297-406), de la promotion (Chapitre 5, p. 407-480), du suivi et de la maintenance (Chapitre 6, p. 481-512) d’un site. Il est complété par des annexes (bibliographie, webographie et table ASCII) et un index. A.M.

(03) CHU, Nicolas. *Réussir un projet de site web*. Paris : Eyrolles, 2003, 212 p. 

Cet ouvrage rédigé par un consultant et responsable de projets informatiques guide le lecteur à travers les étapes d’une conduite de projet web : planification du projet, maîtrise du budget, rédaction d’un cahier des charges, établissement d’une charte graphique et règles de la rédaction pour le Web, élaboration des parcours de navigation et des tests d’utilisabilité, choix raisonné d’un hébergeur, stratégie de référencement, promotion, animation et suivi du site. En annexe sont proposés un modèle de plan projet, de charte graphique et de conditions générales d’utilisation, les critères de référencement des principaux moteurs et annuaires de recherche et un kit comprenant les indicateurs de pilotage d’un projet web.

Le chapitre 4 « Ce qu’est un bon référencement » (p. 95-116) présente une méthodologie pour bien référencer un site. Il aborde successivement les questions d’optimisation des sites web, puis leur soumission aux outils de recherche et les différentes techniques de référencement. Le chapitre 5 (p. 117-145) précise les différentes stratégies de promotion et d’animation (publicité en ligne, affiliation, *netlinking*, cybermarketing, groupes de discussion, communiqués de presse, supports off-line) tandis que le chapitre 6 (p. 147-166) traite exclusivement des analyses et mesures d’audience, lesquels permettent d’identifier le trafic d’un site, de connaître le profil et les comportements des internautes et d’évaluer la notoriété d’un site. A.M.

(04) ANDRIEU, Olivier. Comment fonctionne un moteur de recherche ?. *Abondance.com – Lettre « Recherche & Référencement »*, mars 2005, n° 58, p. 8-18. 

L’auteur propose une analyse globale du fonctionnement des moteurs de recherche et des processus qui sont mis en oeuvre pour traiter les documents, stocker les informations les concernant et présenter des résultats aux recherches des utilisateurs. L’article se poursuit sur une étude plus fine du fonctionnement de Google (architecture, *crawling* et indexation, *ranking*) et s’interroge au final sur les enjeux techniques des moteurs de recherche. A.M.

○ Principe de fonctionnement d'un moteur de recherche

(05) CALISHAIN, Tara et DORNFEST, Rael. *Google à 200% : 100 trucs, secrets et techniques*. Paris : Editions O'Reilly, 2003, 356 p. 

Les auteurs s'adressent principalement aux étudiants, documentalistes et webmasters. Cet ouvrage dévoile 100 trucs, secrets et techniques sur la manière d'effectuer des recherches sur le leader des moteurs, détaille les services proposés, liste les applications non API et API web de Google puis initie le lecteur aux jeux et farces Google. Le chapitre 8 « Google pour les Webmasters » (p. 315-344) nous intéresse plus particulièrement puisqu'il aborde les sujets liés au référencement comme les AdWords, le PageRank ou encore les techniques d'optimisation des pages Web pour Google. A.M.

(06) ANDRIEU, Olivier. 20 milliards de documents pour Yahoo !. *Abondance*, [en ligne], 2005. [référence du 22 août 2005]. <http://actu.abondance.com/2005-32/yahoo-index.php> 

Brève d'actualité se faisant l'écho des récentes déclarations officielles de Yahoo concernant la taille de son index. A.M.

○ Systèmes d'indexation

(07) CACALY, Serge ; LE COADIC, Yves F. ; POMMART, Paul-Dominique et SUTTER Eric. *Dictionnaire de l'information*. Paris : Armand Colin, 2004, 274 p. 

Ce dictionnaire offre près de 600 définitions concernant le vocabulaire technique – voire le jargon – du secteur de l'information/documentation. Il s'intéresse aussi bien aux institutions qu'aux concepts théoriques, outils et méthodes employés dans ce domaine. Il propose en outre des articles de synthèse sur les concepts centraux du domaine et les biographies de personnalités qui ont participé à l'histoire et au développement de l'information. A.M.

(08) AFNOR. *Documentation : Normes, règlements, certification - Tome 1 : Présentation des publications et recherche documentaire*. Paris-La Défense : AFNOR, 2000, 558 p. 

Ce recueil présente les normes relatives à la présentation des publications et à leur numérotation internationale. Il aborde également la translittération, les thésaurus et l'indexation, ou encore l'édition électronique. La norme NF Z 47-102 d'octobre 1993 détaille les principes généraux pour l'indexation des documents : son domaine d'application, son processus et sa finalité, la reconnaissance puis la représentation des concepts dans le langage documentaire, et enfin le contrôle et la qualité de l'indexation. A.M.

(09) CHAUMIER, Jacques. *Travail et méthodes du documentaliste. Pour une exploitation méthodique et optimale de l'information*. Issy-Les-Moulineaux : ESF éd., 2000, 158 p. 

Ce manuel de formation est constitué de trois parties. La première expose des informations théoriques sur différents thèmes (chaîne documentaire, collecte de l'information, analyse documentaire, mémoires et communication documentaires). La seconde partie propose des exercices et leurs corrigés. Les annexes comprennent un plan d'autoformation individuel, les programmes de session de formation, un lexique des termes techniques utilisés et enfin une bibliographie. A.M.

(10) CHIRIE, Pascale. *Les principes de l'indexation*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 1995, 86 p. ✍

Les différents principes d'indexation sont présentés avec leurs limites et avantages respectifs. Tout d'abord, est abordée l'indexation conceptuelle effectuée par les humains fondée sur le modèle booléen. Puis, est traitée, la recherche en texte intégral qui se contente d'apparier des chaînes de caractères et qui ne réclame aucune intervention humaine. Enfin, les méthodes d'indexation automatique ou assistée par ordinateur sont abordées. Les méthodes linguistiques prennent en compte le sens des textes des documents en s'appuyant sur le traitement automatique du langage naturel. Ne sont pas traités ici les outils, les techniques et les produits du marché. *Résumé INTD.*

(11) ANDRIEU, Olivier. *Trouver l'info sur le Web*. Paris : Eyrolles, 2001, 261 p. 📖

Olivier Andrieu est l'un des plus grands spécialistes français du référencement et des outils de recherche sur Internet. Il donne des clés pour formuler des requêtes susceptibles d'améliorer la pertinence des réponses. Il présente également les contextes régissant le choix d'un outil de recherche et ses particularités et aborde rapidement le thème du référencement (p. 232-233). Une bibliographie-webographie, un glossaire et un index viennent clore l'ouvrage. *A.M.*

(12) URIE, Christelle. *Les techniques linguistiques dans les systèmes de recherche d'information – L'exemple de la recherche en langage naturel sur Plutarque, portail d'information en propriété industrielle de l'INPI*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 2004, 136 p. ✍

Le mémoire rappelle, premièrement, les méthodes « traditionnelles » de traitement et recherche de l'information (langages documentaires, texte intégral et syntaxe d'interrogation). Après avoir rappelé leurs limites et l'arrivée, avec l'avènement du texte intégral, de nouveaux usagers, l'auteure met en évidence la nécessité d'introduire des traitements linguistiques. S'attachant tout particulièrement à l'aspect modélisation de la langue et des connaissances, la deuxième partie définit le traitement automatique des langues. Ensuite, les moteurs d'indexation et de recherche sont présentés, et leurs fonctionnalités évoquées à travers la présentation de Lingway KM, moteur sémantique multilingue. Adoptant tour à tour un point de vue documentaire puis un point de vue utilisateur, la troisième partie discute des atouts et limites des moteurs. Enfin, la dernière partie retrace les projets d'intégration d'un moteur de recherche en langage naturel sur Plutarque, portail d'information en propriété industrielle de l'INPI. *Résumé d'auteur.*

○ Méthode de tri des résultats et *ranking*

(13) SULLIVAN, Danny. Search Engine Placement Tips. *Searchenginewatch.com*, [en ligne], 2002. [référence du 27 août 2005].

<http://searchenginewatch.com/webmasters/article.php/2168021>. 🖨

L'auteur définit les critères à prendre en compte et les moyens d'optimiser les pages web pour favoriser le référencement. *A.M.*

(14) AMEROUALI, Youcef. Métadonnées basées sur des éléments de description de ressources et profils d'utilisateur. In : METZGER, Jean-Paul ; HASSOUN, Mohamed ; LAROUK, Omar (éd.), *L'indexation à l'ère d'Internet, Lyon, 21 et 22 octobre 1999*. Lyon : ISKO-France, 2001, p. 43-48. 

Après une introduction aux différents standards de « metadata » existants, l'auteur expose les éléments clés d'un travail de thèse de doctorat. Il propose d'associer à une description des ressources en Dublin Core, des éléments de profil d'utilisateur. A.M.

(15) MOREL-PAIR, Catherine. *Dublin Core et métadonnées : principes, formats, utilisations*. IST, Juin 2004, 141 p. 

Cette présentation Powerpoint fut projetée à l'occasion d'une Rencontre des professionnels de l'IST. Elle détaille les différents formats de métadonnées. A.M.

(16) CRAVEN, Timothy C. Variations in use of meta tag keywords by web pages in different languages. *Journal of Information Science*, 2004, vol. 30, n°3, p. 268-279. 

Les titres et mots-clés de pages rédigées en 19 langues ont été analysés. Les résultats montrent des différences significatives dans le nombre de mots-clés contenus dans une page en fonction de la langue : les pages en hollandais, français et allemand comportent un grand nombre de mots-clés tandis que les pages en chinois et coréen en contiennent fort peu. La proportion de très longs titres varie aussi significativement avec la langue : près de 10 % des titres des pages russes excèdent 100 bytes, contre moins de 1 % sur les pages chinoise, finlandaise, indonésienne et polonaise. A.M.

(17) ANDRIEU, Olivier. Méthodologie d'optimisation d'une page web. *Abondance.com*, [en ligne], 2005, [référence du 18 avril 2005].

<http://methodologies.abondance.com/optimisation.html>. 

Olivier Andrieu dispense ses conseils pour optimiser au mieux une page web. A.M.

(18) DACHELET, Roland. Description des ressources électroniques : quelques éléments. In : INRIA (ed.), *Créer et maintenir un service Web, Pau, 28 septembre-2 octobre 1998*. Paris : ADBS éd., 1998, p. 175-200. 

Dans le cadre de la mise en place des bibliothèques électroniques, l'auteur s'interroge sur le rôle des métadonnées comme moyen de description des ressources électroniques. Il présente quelques exemples significatifs d'élaboration de métadonnées et distingue trois grands formats : Bib-1 issu de l'univers catalographique, IAFA (Internet Anonymous FTP Archive) issu de l'univers d'Internet et enfin un format issu des deux précédents : le Dublin Core. Il s'intéresse également à l'architecture RDF (Resource Description Format) élaborée par le W3C qui permet de gérer les métadonnées. A.M.

(19) BUREAU, Suzanne et EBACHER, Nathalie. Le Dublin Core, un exemple de métadonnées. *Argus*. 2000, vol. 29, n°2, p. 23-30. 

Dans cet article, l'étude d'un cas particulier servira à mieux comprendre le concept de métadonnées. Après une courte introduction sur l'univers des métadonnées en général, les auteures présenteront le Dublin Core, qui est un ensemble d'éléments de métadonnées conçu pour faciliter la découverte des ressources électroniques disséminées sur le Web. Il s'agit d'un système simple de description des ressources pouvant être utilisé autant par des non-catalogueurs que par des spécialistes de la description des ressources. Quelques projets sont également mentionnés, qu'ils soient basés ou non sur le Dublin Core, ainsi que quelques outils qui utilisent le Dublin Core. On constatera que le monde des métadonnées est vaste, mais très intéressant à parcourir. *Résumé d'auteur*.

(20) SULLIVAN, Danny. How Search Engines Rank Web Pages. *Searchenginewatch.com*, [en ligne], 2003. [référence du 27 août 2005].

<http://searchenginewatch.com/webmasters/article.php/2167961>. 

Cet article porte sur les techniques utilisées par les moteurs de recherche pour classer les pages (localisation et fréquence des mots-clés, critères off line). *A.M.*

(21) ANDRIEU, Olivier. L'astuce du mois : quel indice de densité optimum pour vos mots clés ?. *Abondance.com – Lettre « Recherche & Référencement »*, février 2005, n° 57, p. 3-4.



L'indice de densité des mots clés est l'un des critères de pertinence utilisés par les algorithmes des moteurs de recherche. L'auteur propose des astuces pour améliorer la visibilité des sites Internet. *A.M.*

(22) ZHAO, Lisa. Jump Higher : Analyzing Web-site rank in Google. *Information Technology and Libraries*, September 2004, vol. 23, n°3, p. 108-118. 

Compte tenu de l'augmentation exponentielle des sites web et de l'attention très limitée des internautes sur les pages de résultats de recherche des moteurs, il devient nécessaire de comprendre sur quels critères un site web apparaît en tête de liste sur la page de résultats de recherche d'un moteur. Cette étude a consisté à mener une recherche sur Google une fois par semaine pendant 10 semaines. Les facteurs suivants ont été étudiés : le « page rank » (PR) selon Google, la « popularité » des sites web selon Google, le nombre et la densité des mots-clés, leur position dans le titre, dans le corps de la page et dans l'URL et enfin, la structure des noms de domaines. *Résumé Bulletin bibliographique INTD*, vol. 29, n° 7, Juillet 2005, notice 023713, p. 22.

(23) SAPORTA, Sandrine. *Référencement sur le net : les techniques gagnantes*. Paris : Editions d'Organisation, 2005, 187 p. 

Choisir les bonnes techniques de référencement permet d'assurer une présence rentable sur Internet. Cet ouvrage s'adresse avant tout aux responsables marketing qui souhaitent connaître les avantages et inconvénients des différents modes de référencement. Outre un panorama complet des techniques de référencement, des plus traditionnelles (soumission, indexation, positionnement) aux plus récentes (liens sponsorisés et contextuels, *trust feed*), l'auteur propose des fiches pratiques et des exemples concrets pour illustrer chaque problématique. Un lexique des termes techniques et un mémento juridique complètent l'ouvrage. *A.M.*

○ Limites des moteurs de recherche

(24) RAMBAUX-YAKOVLEFF, Nathalie. *Les métadonnées : des normes aux applications concrètes*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 1999, 158 p. ✍

Bilan des modèles de métadonnées et rencontre avec différents acteurs du Web. *Résumé INTD*.

○ La recherche avec un moteur

(25) WEIDEMAN, Melius and STRÜMPFER, Corrie. The effect of search engine keyword choice and demographic features on Internet searching success. *Information Technology and Libraries*, June 2004, p. 58-65. 📖

L'objectif de ce projet était de déterminer les rapports entre le choix des mots-clés et les caractéristiques démographiques d'une population donnée, et le succès de la recherche sur internet. Pour cela, l'expérience a concerné plus de mille étudiants venant de 46 instituts disséminés sur trois continents. Les mots-clés utilisés furent croisés avec leurs âge, race et sexe, et comparés aux résultats des recherches. L'auteur explique comment cette étude a été menée, le travail préalable nécessaire, la méthodologie utilisée, analyse les résultats et en tire conclusion et recommandations. *Bulletin Bibliographique INTD*, vol. 29, n°5, mai 2005, notice 023478, p. 26.

(26) JOURNAL DU NET. France : les outils de recherche. *Journal du Net*, [en ligne], 22 juillet 2005, [référence du 6 septembre 2005].

http://www.journaldunet.com/cc/03_internetmonde/interfrance_moteurs.shtml. 🖨

Cette page regroupe des tableaux statistiques sur les outils de recherche de 2002 à 2005 : utilisation et classement des outils de recherche, typologie des requêtes, part des requêtes images sur le total des requêtes, trafic généré par les outils de recherche. A.M.

(27) CASTAGNE, Rosemarie. *Evolutions et enjeux d'une base documentaire diffusée via Intranet. Comment ajuster les solutions aux besoins des utilisateurs*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 2004, 110 p. ✍

La mise en œuvre d'une base documentaire sur un Intranet ou portail d'entreprise est aujourd'hui amplement facilitée par les solutions techniques existantes. Ainsi, si les technologies ne sont plus un frein, une question reste posée aux professionnels de la gestion d'informations : Quelles fonctionnalités de recherche et quelles interfaces mettre à la disposition des utilisateurs ? Cette étude réalisée lors d'une mission au sein de la CEIDFP, présente des éléments de réflexion et outils méthodologiques pour concevoir l'ouverture d'une BDD au public. Après avoir défini l'Intranet documentaire et ses enjeux, une démarche orientée utilisateur est envisagée prenant appui sur les concepts d'ergonomie cognitive, d'utilisabilité et l'analyse de pratiques d'usage. Deux récits d'expérience et une enquête utilisateur complètent cette approche dans l'objectif d'adapter les solutions (thésaurus en ligne, langages d'indexation, dispositifs intuitifs et personnalisation de services). En dernier lieu, cette étude s'intéresse aux impacts des portails d'entreprise sur le rôle des professionnels et leur implication dans le système d'information (SI). *Résumé d'auteur*.

(28) REFERENCEUR. Référencement – Recommandations pour le choix des mots clés. *Référenceur.com*, [en ligne], [référence du 26 juin 2005]. <http://www.sam-mag.com/P1,1,5,407,,,default.aspx>. 

Referenceur.com présente les six points fondamentaux à respecter pour choisir des mots-clés et assurer la réussite du référencement. A.M.

CHAPITRE 2 - LE REFERENCEMENT : UNE TECHNIQUE DOCUMENTAIRE OU MARKETING ?

○ Intérêt et enjeux du référencement

(29) REVUE-REFERENCEMENT. Panorama du web français. Mai 2005. *Revue-referencement*, [en ligne], mai 2005, [référence du 6 septembre 2005]. http://www.revue-referencement.com/CHIFFRESCLES/panorama_mai05.htm. 

Tableaux basés sur des statistiques e-stat/@position de mai 2005 : classement des outils de recherche, détails des marques Google et Yahoo, répartition du trafic entrant par type d'accès. A.M.

(30) XUE, Susan. Web usage statistics and Web site evaluation : a case study of a government publications library Web site. *Online Information Review*, 2004, vol. 28, n°3, p. 180-190. 

Les documents officiels émanant d'agences gouvernementales sont très nombreux aux Etats-Unis. Donner accès à toutes les agences et à leurs documents en ligne à partir d'un site web de bibliothèque demande de réfléchir à l'organisation et à la catégorisation de l'information. Cette étude vise à évaluer le site web d'une bibliothèque offrant ce type d'accès à partir de ses statistiques d'utilisation. Cette méthode prouve l'intérêt d'offrir une arborescence par thématiques mais il faut sans cesse la développer. Elle permet enfin de mieux gérer le référencement des documents sur les moteurs de recherche. *Bulletin bibliographique INTD*, vol. 29, n°2, février 2005, notice 023165, p.46.

(31) BOULOGNE, Arlette (coord.). *Vocabulaire de la documentation*. Paris : ADBS, 2004, 334 p. 

Destiné aux professionnels de l'information-documentation, cet ouvrage dévoile le vocabulaire technique de leur activité à travers la définition de 685 termes. Les notices sont classées alphabétiquement. Chacune comprend, avant la définition du terme considéré, des renvois vers des termes apparentés à celui-ci, la mention du ou des domaines auxquels il se rattache et sa traduction en anglais, allemand et espagnol. Une liste des termes définis et trois tables de correspondance renvoyant aux traductions complètent l'ouvrage. A.M.

○ Optimiser une page web pour le référencement

(32) QUATRAVAUX, Anne-Laure et QUATRAVAUX, Dominique. *Réussir un site web d'association...avec des outils gratuits !*. Paris : Eyrolles, 2004, 334 p. 

Destiné avant tout aux webmasters et créateurs de sites web associatifs, cet ouvrage rédigé à deux mains veut aider les associations (souvent handicapées par de faibles ressources) à améliorer leur gestion et leur visibilité sur le Web avec des outils gratuits. Le chapitre 10 (p. 239-261), axé sur le référencement, la promotion et le suivi d'un site web, retiendra plus particulièrement notre attention. A.M.

(33) LE GUELVOUIT, Arnaud. *Comment concevoir un service web : de la théorie à la pratique*. Paris : ADBS Edition, 1999, 147 p. 

L'auteur présente les fondements de l'Internet et du web puis les étapes successives d'un projet de service web depuis sa conception jusqu'à sa réalisation avant d'aborder sa maintenance. Cette dernière, qui fait l'objet du sixième chapitre (p. 122-136), s'appuie sur différentes démarches : analyse des connexions, référencement sur les outils de recherche, promotion et évaluation du service. Une liste de liens, pour la majorité périmés, et une brève bibliographie, viennent clore le chapitre. A.M.

○ Un référencement différencié en fonction des outils de recherche et des objectifs

(34) FOENIX-RIOU, Béatrice. Référencement payant : les plus riches seront les premiers !. *Bases/Netsources*. Mai/Juin 2002, n° 38, p. 10-13. 

Ce compte-rendu de la journée organisée à Lyon le 28 mai 2002 par l'association Doc Forum et Abondance sur le thème "Référencement, vers le tout payant ?" se fait l'écho des points de vue des référenceurs et des outils de recherche sur cette question. L'article explique le succès du référencement payant puis présente le paysage du référencement en 2002 et les types d'offres disponibles (soumissionnement, indexation et positionnement payants) avant de conclure sur leurs conséquences pour l'utilisateur. A.M.

(35) ZHANG, Jin et DIMITROFF, Alexandra. Internet Search Engine's response to metadata Dublin Core implementation. *Journal of Information Science*. 2004, vol. 30, n° 4, p. 310-320.



L'étude décrite dans cet article s'intéresse à la performance des principaux moteurs de recherche par rapport à deux groupes de pages web : celles contenant des métadonnées et celles qui n'en contiennent pas. Le champ "subject" du Dublin Core peut améliorer l'accessibilité d'une page web et, de fait, sa visibilité. Les résultats montrent qu'il existe des différences significatives dans le traitement des deux groupes de pages web par de nombreux moteurs. Pour Altavista, en revanche, les métadonnées ont peu d'impact sur la visibilité des pages. Les raisons qui expliquent ces différences sont analysées. A.M.

(36) LE GUELVOUIT, Arnaud. Les outils de recherche du web : Limites et aléas du référencement. *Documentaliste-Sciences de l'information*, 1998, vol. 35, n° 6, p. 315-320. 

Cet article s'interroge sur l'impact et l'efficacité des métadonnées à partir de l'exemple du site web développé en 1998 par l'INRIA pour présenter son cours sur les nouvelles technologies dans l'IST : présentation des modalités d'indexation puis comparaison du positionnement du site sur différents outils francophones de recherche sur le web. A.M.

(37) CARLIER, Richard. Accueillir de la pub sur son site. *Richardcarlier.com*, [en ligne], 1 mars 2005, [référence du 5 septembre 2005]. <http://www.richardcarlier.com/accueillir-pub-son-site.php>. 

Ce court article détaille le principe de fonctionnement général de la publicité sur Internet, le positionnement payant sur les mots-clés, les dérives et risques que cela comporte et enfin les avantages des programmes de publicité pour l'éditeur. A.M.

(38) MESLON, Julie (de). Les liens sponsorisés, l'Eldorado du Web. *Décision Informatique*, [en ligne], 19 août 2005, [référence du 23 août 2005]. <http://www.01net.com/article/286163.html>. 

Liens sponsorisés et autres (liens contextuels) rencontrent un vrai succès. Les raisons qui expliquent ce phénomène sont clairement indiquées dans cet article. A.M.

(39) ANDRIEU, Olivier. Quel avenir pour le "Paid inclusion". *Abondance.com – Lettre « Recherche & Référencement »*, 2005, n° 56, p. 11-19. 

Tandis que trois des quatre principaux moteurs de recherche (Google, MSN et Ask Jeeves) ne disposent d'aucune offre de "paid inclusion", on peut s'interroger sur l'avenir de cette méthode de référencement payant. Yahoo ! fait figure de leader sur ce segment et de nombreux autres moteurs nationaux, de taille parfois plus modeste, proposent toujours une telle offre, faisant en particulier valoir l'intérêt de ces solutions pour le référencement des sites dynamiques. Alors, faut-il oui ou non prendre en compte ces offres et pour quels types de contenus ?. *Résumé d'auteur*.

(40) ANDRIEU, Olivier. "Paid inclusion" : l'avis des référenceurs. *Abondance.com – Lettre « Recherche & Référencement »*, 2005, n° 56, p. 20-23. 

L'auteur a posé quatre questions à plusieurs spécialistes du référencement concernant leur utilisation des programmes de "Paid inclusion", les avantages et inconvénients de ces programmes, en particulier pour le référencement et leur vision sur son avenir, que la majorité d'entre eux juge plutôt sombre. A.M.

(41) ADCOM. Référencement des sites dynamiques. *Adcom*, [en ligne], 2003, [référence du 31 août 2005]. <http://www.adcom.fr/expertise/referencement.asp.htm>. 

Cet article rédigé par une société professionnelle de référencement indique quelles techniques facilitent le référencement des pages dynamiques sur les moteurs de recherche. A.M.

○ Sociétés spécialisées dans le référencement

(42) QUIN, Frédéric. Quand le référencement évolue vers le *search engine marketing*, *Le Journal du Net*, [en ligne], 2004, [référence du 8 juin 2005]. <http://www.journaldunet.com/0409/040909referencement.shtml>. 

Cet article retrace l'évolution des offres des référenceurs pour répondre aux besoins des clients. A.M.

(43) ANDRIEU, Olivier. Résultats d'une étude sur le référencement et les liens sponsorisés. *Abondance.com – Lettre « Recherche & Référencement »*, février 2005, n° 57, p. 23-32. 

Résultats d'une enquête menée en février 2005 auprès de plus de 500 personnes sur le référencement de sites web et les liens sponsorisés. A.M.

CHAPITRE 3 - LE REFERENCEMENT : UNE ETAPE DANS LE PLAN MARKETING D'UN SITE WEB

○ Le site web : un produit

(44) MUET, Florence et SALAÜN, Jean-Michel. *Stratégie marketing des services d'information, bibliothèques et centres de documentation*. Paris : Editions du Cercle de la Librairie, 2001, 221 p. 

[...] Suivant une approche marketing, ce manuel détaille la démarche stratégique indispensable au développement des services d'information [...]. Chaque chapitre développe une étape de la phase d'analyse et de diagnostic puis de la phase de stratégie et se conclut par un outil directement transposable sur le terrain. Trois études de cas illustrent en fin de volume la faisabilité et l'intérêt de la méthode [...]. *Résumé d'auteur*.

(45) BERNADET, J.-P., BOUCHEZ, A., PIHIER, S. *Précis de marketing*. Paris : Nathan, 2001, 159 p. 

Ce petit ouvrage présente de façon synthétique les thèmes liés au marketing. Divisé en six parties (fondements, analyse de marché, produit, distribution, publicité, communication), il permet d'embrasser l'ensemble de la sphère marketing à travers des pages de synthèse apportant des informations essentielles pour comprendre un sujet et des pages d'explication qui développent un point particulier venant illustrer et compléter la page de synthèse. A.M.

○ Les préalables au référencement

(46) REFERENCEUR. Choix des mots clés et référencement : une question d'adéquation mots clés contenu !. *Référenceur.com*, [en ligne], [référence du 26 juin 2005].

<http://www.sam-mag.com/P1,1,5,406,,,default.aspx>. 

Ce bref article proposé par une société spécialisée dans le référencement donne des conseils pour élaborer une liste pertinente de mots-clés représentative du contenu. Il aborde successivement le poids des mots-clés dans le fonctionnement des moteurs de recherche, les moyens de déterminer puis de contrôler la qualité des mots-clés retenus. *A.M.*

(47) MUET, Florence. Positionnement du site web documentaire. In : INRIA (ed.), *Créer et maintenir un service Web,, Pau, 28 septembre-2 octobre 1998*. Paris : ADBS éd., 1998, p. 13-41. 

L'Institut National de Recherche en Informatique et en Automatique (INRIA) publie ici le contenu des cours qu'il a organisé sur la mise en place et la maintenance d'un service web. Dans cette contribution, Florence Muet applique l'approche marketing documentaire à la création d'un site web. Elle détaille l'environnement marketing, les différents scénarii de positionnement de l'offre et dresse un plan marketing pour le site. *A.M.*

(48) GUENTHER, Kim. Getting your web site recognized. *Online Magazine*. May/June 2004, p. 47-49. 

L'auteur, directrice du *Health System Web Center* de l'université de Virginie et webmaster pour le système de santé UVA, dévoile ici les stratégies nécessaires à l'optimisation des pages web en vue de leur indexation dans les moteurs de recherche. Elle présente succinctement le mode de fonctionnement des moteurs de recherche, regrette l'absence de standard d'indexation avant de replacer la question dans le cadre plus vaste du champ des connaissances et de rappeler les stratégies de promotion et d'animation d'un site web. *A.M.*

○ Analyses et mesures d'audience

(49) ADCOM. Fichier log. *Adcom*, [en ligne], 2003, [référence du 31 août 2005].

http://www.adcom.fr/expertise/fichier_log.htm. 

La société de référencement Adcom répond brièvement aux questions que l'on peut se poser sur les fichiers log : comment se présentent-ils, que peut-on y trouver, comment les analyser ? Une suggestion de sites proposant des logiciels d'analyse est fournie à la fin de l'article. *A.M.*

(50) ADCOM. Le trafic. *Adcom*, [en ligne], 2003, [référence du 31 août 2005].

<http://www.adcom.fr/expertise/trafic.htm>. 

Ce court article expose les moyens de quantifier et d'évaluer le trafic d'un site web. *A.M.*

(51) BERANGER, Anne-Laure. Comment ne pas rater son référencement, *Le Journal du Net*, [en ligne], 21 février 2005, [référence du 8 juin 2005]. <http://www.journaldunet.com/0502/050221referencement.shtml> 

Bien référencer un site est une opération stratégique de laquelle peut dépendre 30 à 40 % du trafic. Dix conseils qui peuvent faire la différence. *Résumé d'auteur*.

○ Le référencement : une méthode parmi d'autres dans le processus de promotion d'un site web

(52) ANDRIEU, Olivier. Les fils RSS, le référencement et la recherche d'information. *Abondance.com-Lettre « Recherche & Référencement »*. Décembre 2004, n° 55, p. 13-24. 

Les fils ou flux RSS sont de plus en plus nombreux. Ils permettent de rester informés des dernières modifications effectuées sur un site ou sur un service d'information. Envoyés au destinataire en mode « push », ils sont intégrés sous forme de liens cliquables dans des logiciels spécialisés que l'on appelle « agrégateur » ou « lecteur RSS » et qui convertissent le XML en HTML. Cet article propose une typologie des contenus des fils RSS ainsi que des tendances d'avenir ; une typologie techniques des lecteurs RSS sous la forme d'un tableau comparatif ; quelques outils de recherche de fils RSS et une liste de fils intéressants. Les intérêts des fils RSS sont détaillés, notamment pour les référenceurs. *Résumé Bulletin bibliographique INTD vol. 29, n° 5, mai 2005, notice 023524*.

(53) CROWDER, David et CROWDER, Rhonda. *Créer un site web pour les nuls*. Paris : Editions First Interactive, 2003, 263 p. 

Après quelques généralités sur les sites web, les auteurs abordent les moyens d'animer un site ou de communiquer avec les visiteurs. Ils dévoilent au lecteur les caractéristiques des logiciels WYSIWYG, puis s'intéressent à l'hébergement, au référencement et à la promotion du site. Les adresses Internet fournies pour compléter ou illustrer les propos sont pour beaucoup périmées. Sur le sujet qui nous intéresse plus particulièrement, nous renvoyons les lecteurs au chapitre 13 « Faire connaître votre site » (p. 183-196) qui explique comment bien choisir ses mots-clés, se référencer, vérifier sa position, échanger des liens ou utiliser des bandeaux publicitaires. Un glossaire et un index complètent le tout. *A.M.*

CHAPITRE 4 – ETUDE COMPARATIVE : LES SITES WEB DES CYBERLIBRAIRIES ET MAISONS D'EDITION

○ Présentation des sociétés et sites étudiés

(54) MOULIN, Olivia. Dix ans de règne sur le commerce en ligne mondial. In : *Le Monde*, 17 juillet 2005. 

Cet article retrace la genèse puis l'évolution de la plus grosse entreprise de e-commerce du monde. *A.M.*

(55) KARAYAN, Raphaële. Trois questions à... Xavier Garambois (Amazon France) : « La France reste un marché important pour le groupe ». *JNet Solutions*, [en ligne], 28 juillet 2005, [référence du 14 août 2005]. <http://www.journaldunet.com/0507/050728amazon.shtml>.



L'auteur interroge l'actuel dirigeant de la filiale française d'Amazon. Ce dernier fait part de sa confiance dans le potentiel de développement de sa société. A.M.

(56) COMMENT CA MARCHE, Introduction aux Active Server Pages (ASP), *Comment ça marche*, [en ligne], [référence du 15 septembre 2005]. <http://www.commentcamarche.net/asp/aspintro.php3>

La célèbre encyclopédie informatique basée sur la licence Creative Commons explique simplement ce que sont l'ASP et les objets ASP. Les rubriques concernant l'implantation du code et la configuration s'adressent davantage aux initiés. A.M.

(57) COMMENT CA MARCHE, Introduction à PHP, *Comment ça marche*, [en ligne], [référence du 15 septembre 2005]. <http://www.commentcamarche.net/php/phpintro.php3>

Le PHP est un langage de script largement utilisé dans les sites dynamiques sur le net. Comment ça marche nous conte les origines du PHP puis détaille ses caractéristiques. L'implantation du code, l'installation (sous Linux ou Windows) et le guide utilisateur permettent aux intéressés d'approfondir leurs connaissances. A.M.

(58) CLAPAUD, Alain. Alapage déploie ses services web avec PHP. *01net*, [en ligne], 1^{er} avril 2005, [référence du 6 septembre 2005]. <http://www.01net.com/article/275940.html>

Le site alapage.com dynamise sa politique d'affiliation et dévoile son offre de services web. Ceux-ci mettent à profit la plate-forme PHP exploitée par le cybermarchand. *Résumé d'auteur*.

(59) DIZ, José. Les dessous techniques d'Amazon.fr. *JNet Solutions*, [en ligne], 2000, [référence du 14 août 2005]. <http://solutions.journaldunet.com/0008/000830amazontech.shtml>



Cet article, qui date un peu, est, avec celui de Simottel (voir ci-dessous), l'un des rares à exposer les solutions techniques mises en place par Amazon pour faire fonctionner son site de e-commerce. A.M.

(60) SIMOTTEL, Frédéric. Les dessous techniques de l'arrivée d'Amazon.fr. *Internet Professionnel*, [en ligne], 1^{er} octobre 2000, [référence du 14 août 2005] <http://www.01net.com/article/127582.html>

En 2000, Amazon hébergeait tous ses serveurs aux États-Unis. L'article présente rapidement la stratégie technique prudente de la filiale française. A.M.

○ Optimisation des pages

(61) SOLOVIEFF, Karine. Fnac.com, Amazon.fr : le conseil pour l'un l'efficacité pour l'autre. *01net*, [en ligne], 13 février 2003, [référence du 23 août 2005].

<http://www.01net.com/article/201808.html> 

Selon une étude basée sur le comportement des internautes, Amazon.fr attire plus d'acheteurs que Fnac.com. Mais le site de la Fnac retient plus longtemps les visiteurs. *Résumé d'auteur*.

(62) ADCOM, Les balises méta. *Adcom*, [en ligne], 2003, [référence du 31 août 2005].

<http://www.adcom.fr/expertise/balisesmeta.htm> 

Adcom, société spécialisée en référencement et promotion de site web explique succinctement l'intérêt et l'optimisation des balises meta. *A.M.*

Bibliographie alphabétique

(classement par noms d'auteurs)

🖨 **ADCOM.** Fichier log. *Adcom*, [en ligne], 2003, [référence du 31 août 2005].
http://www.adcom.fr/expertise/fichier_log.htm.

🖨 **ADCOM,** Les balises méta. *Adcom*, [en ligne], 2003, [référence du 31 août 2005].
<http://www.adcom.fr/expertise/balisesmeta.htm>.

🖨 **ADCOM.** Le trafic. *Adcom*, [en ligne], 2003, [référence du 31 août 2005].
<http://www.adcom.fr/expertise/trafic.htm>.

🖨 **ADCOM.** Référencement des sites dynamiques. *Adcom*, [en ligne], 2003, [référence du 31 août 2005]. http://www.adcom.fr/expertise/referencement_asp.htm.

📖 **AFNOR.** *Documentation : Normes, règlements, certification - Tome 1 : Présentation des publications et recherche documentaire.* Paris-La Défense : AFNOR, 2000, 558 p.

📖 **AMAR, Muriel.** *Les fondements théoriques de l'indexation : une approche linguistique.* Paris : ADBS éd., 2000, 355 p.

📖 **AMEROUALI, Youcef.** Métadonnées basées sur des éléments de description de ressources et profils d'utilisateur. In : METZGER, Jean-Paul ; HASSOUN, Mohamed ; LAROUC, Omar (éd.), *L'indexation à l'ère d'Internet, Lyon, 21 et 22 octobre 1999.* Lyon : ISKO-France, 2001, p. 43-48.

🖨 **ANDRIEU, Olivier.** 20 milliards de documents pour Yahoo !. *Abondance*, [en ligne], 2005. [référence du 22 août 2005]. <http://actu.abondance.com/2005-32/yahoo-index.php>.

📄 **ANDRIEU, Olivier.** Comment fonctionne un moteur de recherche ?. *Abondance.com* – Lettre « Recherche & Référencement », mars 2005, n° 58, p. 8-18.

📄 **ANDRIEU, Olivier.** L’astuce du mois : quel indice de densité optimum pour vos mots clés ?. *Abondance.com* – Lettre « Recherche & Référencement », février 2005, n° 57, p. 3-4.

📄 **ANDRIEU, Olivier.** Résultats d’une étude sur le référencement et les liens sponsorisés. *Abondance.com* – Lettre « Recherche & Référencement », février 2005, n° 57, p. 23-32.

📄 **ANDRIEU, Olivier.** "Paid inclusion" : l’avis des référenceurs. *Abondance.com* – Lettre « Recherche & Référencement », janvier 2005, n° 56, p. 20-23.

📄 **ANDRIEU, Olivier.** Quel avenir pour le "Paid inclusion" ?. *Abondance.com* – Lettre « Recherche & Référencement », janvier 2005, n° 56, p. 11-19.

📄 **ANDRIEU, Olivier.** Méthodologie d’optimisation d’une page web. *Abondance.com*, [en ligne], 2005, [référence du 18 avril 2005].
<http://methodologies.abondance.com/optimisation.html>.

📄 **ANDRIEU, Olivier.** Les fils RSS, le référencement et la recherche d’information. *Abondance.com* - Lettre « Recherche & Référencement », décembre 2004, n° 55, p. 13-24.

📖 **ANDRIEU, Olivier.** *Trouver l’info sur le Web*. Paris : Eyrolles, 2001, 261 p.

📖 **ANDRIEU, Olivier.** *Créer du trafic sur son site web : conception, référencement, promotion*. Paris : Eyrolles, 2000, 532 p.

📄 **BERANGER, Anne-Laure.** Comment ne pas rater son référencement, *Le Journal du Net*, [en ligne], 21 février 2005, [référence du 8 juin 2005].
<http://www.journaldunet.com/0502/050221referencement.shtml>.

📖 **BERNADET, J.-P., BOUCHEZ, A., PIHIER, S.** *Précis de marketing*. Paris : Nathan, 2001, 159 p.

📖 **BOULOGNE, Arlette** (coord.). *Vocabulaire de la documentation*. Paris : ADBS, 2004.- 334 p.

📄 **BUREAU, Suzanne** et **EBACHER, Nathalie**. Le Dublin Core, un exemple de métadonnées. *Argus*. 2000, vol. 29, n°2, p. 23-30.

📖 **CACALY, Serge (dir.)**. *Dictionnaire de l'information*. Paris : Armand Colin, 2004, 274 p.

📖 **CALISHAIN, Tara** et **DORNFEST, Rael**. *Google à 200% : 100 trucs, secrets et techniques*. Paris : Editions O'Reilly, 2003, 356 p.

💻 **CARLIER, Richard**. Accueillir de la pub sur son site. *Richardcarlier.com*, [en ligne], 1 mars 2005, [référence du 5 septembre 2005]. <http://www.richardcarlier.com/accueillir-pub-son-site.php>.

📖 **CARLIER, Richard**. *Profession webmaster*. Paris : « Les Echos » Ed., 2000, 255 p.

✍️ **CASTAGNE, Rosemarie**. *Evolutions et enjeux d'une base documentaire diffusée via Intranet. Comment ajuster les solutions aux besoins des utilisateurs*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 2004, 110 p.

📖 **CHAUMIER, Jacques**. *Travail et méthodes du documentaliste. Pour une exploitation méthodique et optimale de l'information*. Issy-Les-Moulineaux : ESF éd., 2000, 158 p.

✍ **CHIRIE, Pascale.** *Les principes de l'indexation.* Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 1995, 86 p.

📖 **CHU, Nicolas.** *Réussir un projet de site web.* Paris : Eyrolles, 2003, 212 p.

💻 **CLAPAUD, Alain.** Alapage déploie ses services web avec PHP. *01net*, [en ligne], 1^{er} avril 2005, [référence du 6 septembre 2005]. <http://www.01net.com/article/275940.html>

💻 **COMMENT CA MARCHE,** Introduction à PHP, *Comment ça marche*, [en ligne], 2005, [référence du 15 septembre 2005]. <http://www.commentcamarche.net/php/phpintro.php3>

💻 **COMMENT CA MARCHE,** Introduction aux Active Server Pages (ASP), *Comment ça marche*, [en ligne], 2005 [référence du 15 septembre 2005]. <http://www.commentcamarche.net/asp/aspintro.php3>

📄 **CRAVEN, Timothy C.** Variations in use of meta tag keywords by web pages in different languages. *Journal of Information Science*, 2004, vol. 30, n°3, p. 268-279.

📖 **CROWDER, David** et **CROWDER, Rhonda.** *Créer un site web pour les nuls.* Paris : Editions First Interactive, 2003, 263 p.

📄 **DACHELET, Roland.** Description des ressources électroniques : quelques éléments. In : INRIA (ed.), *Créer et maintenir un service Web, Pau, 28 septembre-2 octobre 1998.* Paris : ADBS éd., 1998, p. 175-200.

💻 **DIZ, José.** Les dessous techniques d'Amazon.fr. *JDNet Solutions*, [en ligne], 2000, [référence du 14 août 2005]. <http://solutions.journaldunet.com/0008/000830amazontech.shtml>.

📄 **FOENIX-RIOU, Béatrice.** Référencement payant : les plus riches seront les premiers !. *Bases/Netsources*. Mai/Juin 2002, n° 38, p. 10-13.

📄 **GUENTHER, Kim.** Getting your web site recognized. *Online Magazine*. May/June 2004, p. 47-49.

📄 **JOURNAL DU NET.** France : les outils de recherche. *Journal du Net*, [en ligne], 22 juillet 2005, [référence du 6 septembre 2005].
http://www.journaldunet.com/cc/03_internetmonde/interfrance_moteurs.shtml.

📄 **KARAYAN, Raphaële.** Trois questions à... Xavier Garambois (Amazon France) : « La France reste un marché important pour le groupe ». *JDNet Solutions*, [en ligne], 28 juillet 2005, [référence du 14 août 2005]. <http://www.journaldunet.com/0507/050728amazon.shtml>.

📖 **LE GUELVOUIT, Arnaud.** Comment concevoir un service web : de la théorie à la pratique. Paris : ADBS Edition, 1999, 147 p.

📄 **LE GUELVOUIT, Arnaud.** Les outils de recherche du web : Limites et aléas du référencement. *Documentaliste-Sciences de l'information*, 1998, vol. 35, n° 6, p. 315-320.

📄 **MESLON, Julie (de).** Les liens sponsorisés, l'Eldorado du Web. *Décision Informatique*, [en ligne], 19 août 2005, [référence du 23 août 2005].
<http://www.01net.com/article/286163.html>.

📄 **MOREL-PAIR, Catherine.** *Dublin Core et métadonnées : principes, formats, utilisations*, IST [présentation Powerpoint], Juin 2004, 141 p.

📄 **MOULIN, Olivia.** Dix ans de règne sur le commerce en ligne mondial. In : *Le Monde*, 17 juillet 2005.

📖 **MUET, Florence** et **SALAÜN, Jean-Michel**. *Stratégie marketing des services d'information, bibliothèques et centres de documentation*. Paris : Editions du Cercle de la Librairie, 2001, 221 p.

📄 **MUET, Florence**. Positionnement du site web documentaire. In : INRIA (ed.), *Créer et maintenir un service Web, Pau, 28 septembre-2 octobre 1998*. Paris : ADBS éd., 1998, p. 13-41.

📖 **QUATRAVAUX, Anne-Laure** et **QUATRAVAUX, Dominique**. *Réussir un site web d'association...avec des outils gratuits !*. Paris : Eyrolles, 2004, 334 p.

💻 **QUIN, Frédéric**. Quand le référencement évolue vers le *search engine marketing*, *Le Journal du Net*, [en ligne], 2004, [référence du 8 juin 2005].
<http://www.journaldunet.com/0409/040909referencement.shtml>.

✍ **RAMBAUX-YAKOVLEFF, Nathalie**. *Les métadonnées : des normes aux applications concrètes*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 1999, 158 p.

💻 **REFERENCEUR**. Choix des mots clés et référencement : une question d'adéquation mots clés contenu !. *Référenceur.com*, [en ligne], [référence du 26 juin 2005].
<http://www.sam-mag.com/P1,1,5,406,,,default.aspx>.

💻 **REFERENCEUR**. Référencement – Recommandations pour le choix des mots clés. *Référenceur.com*, [en ligne], [référence du 26 juin 2005].
<http://www.sam-mag.com/P1,1,5,407,,,default.aspx>.

💻 **REVUE-REFERENCEMENT**. Panorama du web français. Mai 2005. *Revue-referencement*, [en ligne], mai 2005, [référence du 6 septembre 2005].
http://www.revue-referencement.com/CHIFFRESCLES/panorama_mai05.htm

📖 **SALAÛN, Jean-Michel.** *Marketing des bibliothèques et des centres de documentation.* Paris : Editions du Cercle de la Librairie, 1992, 132 p.

📖 **SAPORTA, Sandrine.** *Référencement sur le net : les techniques gagnantes.* Paris : Editions d'Organisation, 2005, 187 p.

💻 **SIMOTTEL, Frédéric.** Les dessous techniques de l'arrivée d'Amazon.fr. *Internet Professionnel*, [en ligne], 1^{er} octobre 2000, [référence du 6 septembre 2005].
<http://www.01net.com/article/127582.html>.

💻 **SOLOVIEFF, Karine.** Fnac.com, Amazon.fr : le conseil pour l'un l'efficacité pour l'autre. *01net*, [en ligne], 13 février 2003, [référence du 23 août 2005].
<http://www.01net.com/article/201808.html>.

💻 **SULLIVAN, Danny.** How Search Engines Rank Web Pages. *Searchenginewatch.com*, [en ligne], 2003. [référence du 27 août 2005].
<http://searchenginewatch.com/webmasters/article.php/2167961>.

💻 **SULLIVAN, Danny.** Search Engine Placement Tips. *Searchenginewatch.com*, [en ligne], 2002. [référence du 27 août 2005].
<http://searchenginewatch.com/webmasters/article.php/2168021>.

✍ **URIE, Christelle.** *Les techniques linguistiques dans les systèmes de recherche d'information – L'exemple de la recherche en langage naturel sur Plutarque, portail d'information en propriété industrielle de l'INPI.* Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 2004, 136 p.

📄 **WEIDEMAN, Melius and STRÜMPFER, Corrie.** The effect of search engine keyword choice and demographic features on Internet searching success. *Information Technology and Libraries*, June 2004, p. 58-65.

📄 **XUE, Susan.** Web usage statistics and Web site evaluation : a case study of a government publications library Web site. *Online Information Review*, 2004, vol. 28, n°3, p. 180-190.

📄 **ZHANG, Jin** et **DIMITROFF, Alexandra.** Internet Search Engine's response to metadata Dublin Core implementation. *Journal of Information Science*. 2004, vol. 30, n° 4, p. 310-320.

📄 **ZHAO, Lisa.** Jump Higher : Analyzing Web-site rank in Google. *Information Technology and Libraries*, September 2004, vol. 23, n°3, p. 108-118.

Bibliographie complémentaire

📖 AFNOR. *Vocabulaire de la documentation*. Paris-La Défense : AFNOR, 1987, 159 p.

Ce recueil définit les termes en usage et les recommandations en information-documentation. *A.M.*

📄 AMAR, Muriel. L'indexation : un choix de mots ou un choix de textes ? Vers une « indexation discursive » sur Internet. In : METZGER, Jean-Paul ; HASSOUN, Mohamed et LAROUK, Omar (éd.), *L'indexation à l'ère d'Internet, Lyon, 21 et 22 octobre 1999*, Lyon : ISKO-France, 2001, p. 51-56.

Cette communication se propose de poser les enjeux de l'indexation sur Internet en exploitant l'opposition établie en linguistique entre lexique et discours. Cette opposition permet de distinguer deux types d'indexation – lexicale et discursive – qui ne peuvent être pratiquées indifféremment : chacun relève d'un contexte documentaire particulier qu'il importe de caractériser. *Résumé d'auteur.*

📖 AMAR, Muriel. *Les fondements théoriques de l'indexation : une approche linguistique*. Paris : ADBS éd., 2000, 355 p.

L'indexation documentaire est le plus souvent appréhendée par les professionnels de l'information dans sa seule dimension instrumentale : c'est en se focalisant sur sa finalité - la recherche d'information - qu'elle se définit d'abord. Si le choix d'un tel angle d'approche se comprend aisément dans le cadre des pratiques professionnelles, il se révèle réducteur dès lors que l'indexation devient l'objet d'autres problématiques, notamment celles de l'évolution des technologies de l'information.

La recherche d'où est issu cet ouvrage s'est intéressée à dégager les spécificités, les caractéristiques, les propriétés de l'indexation et à en étudier les fondements théoriques par le biais d'interrogations issues de la linguistique. La première partie est consacrée aux "Problèmes théoriques de l'indexation" : la question du lexique et celle de la référence y sont examinées. La seconde partie est une "Contribution aux fondements théoriques de l'indexation" : elle s'attache à faire émerger les aspects "discursifs" de l'indexation, tant au niveau du processus que du résultat. Cette approche permet de situer l'indexation dans le cadre plus large des pratiques de diffusion des connaissances, la rapprochant des problématiques de la vulgarisation scientifique. *Résumé ADBS.*

📄 ANDRIEU, Olivier. L'avenir du référencement sur les outils de recherche. *Bases/Netsources*. Juillet/Août 2001, n° 33, p. 10-12.

Cet article décrit l'évolution prévisible du marché du référencement en 2001/2002. Olivier Andrieu prévoit déjà la généralisation du soumissionnement payant sur les annuaires et du référencement et positionnement payant sur les moteurs de recherche, une tendance qui s'est affirmée depuis. *A.M.*

📄 ANDRIEU, Olivier. Référencement : les gestes qui sauvent. *Archimag*. Mai 2001, n° 144, p. 32-33.

L'auteur fournit des conseils pour optimiser dès le départ un site web en vue de son référencement. Il passe en revue les critères de sélection retenus par les annuaires puis ceux des moteurs de recherche. Il détaille enfin les erreurs à éviter pour qu'un site web puisse être correctement référencé. *A.M.*

📖 BONNET, Delphine. *Créer son site web 2002*. Paris : L'Etudiant, 2002, 217 p.

Delphine Bonnet est journaliste. Elle a conçu cet ouvrage comme un guide de conception de site web. A ce titre, il reprend les règles à respecter pour réaliser un site harmonieux. L'auteur s'attarde sur l'utilisation de Front Page et de Dreamweaver et n'évade pas les indispensables questions sur l'hébergement, la mise à jour et le référencement (p. 205-213) d'un site. Un glossaire et un index complètent cet ouvrage. A.M.

📖 CACALY, Serge (sous la dir. de). Dictionnaire encyclopédique de l'information et de la documentation. Paris : Nathan, 1997, 634 p.

Destiné avant tout aux professionnels et étudiants en information-documentation, ce dictionnaire encyclopédique francophone est dédié aux sciences de l'information et de la documentation. Il est rédigé par près de cent spécialistes et comprend 307 articles qui définissent les concepts fondamentaux du secteur de l'information-documentation (métiers, formations, économie...). A.M.

📖 DAVID, Claire ; GIROUX, Luc ; BERTRAND-GASTALDY, Suzanne. Explication cognitive au problème de la cohérence entre indexeurs. In : METZGER, Jean-Paul ; HASSOUN, Mohamed ; LAROUK, Omar (éd.). *L'indexation à l'ère d'Internet*, Lyon, 21 et 22 octobre 1999. Lyon : ISKO-France, 2001, p. 163-169.

Cette contribution examine le travail intellectuel des indexeurs dans le but de fournir une explication cognitive à leur incohérence, et ceci grâce aux théories et méthodes des sciences cognitives. Selon l'hypothèse retenue, cette incohérence s'expliquerait par la différence issue de la représentation personnelle de chaque indexeur face au processus de l'indexation. L'expérimentation passe par l'analyse de protocoles verbaux. A.M.

📖 DEGEZ, Danièle ; MENILLET, Dominique. *Thésauroglossaire des langages documentaires : un outil de contrôle sémantique*. Paris : ADBS, 2001, 185 p. (Sciences de l'information, série Recherches et documents).

Ce "thésauroglossaire" est un outil terminologique relatif au domaine de l'analyse, de l'indexation et des langages documentaires.

Il propose un "glossaire", liste alphabétique de définitions des termes appartenant au domaine étudié, qui proviennent de l'univers des thésaurus comme de celui des classifications ; et un "thésaurus" qui replace tous ces termes dans leur environnement sémantique et met en évidence les relations qu'ils entretiennent entre eux. De nombreux exemples de langages documentaires illustrent ce thésauroglossaire conçu par Danièle Dégez et Dominique Ménillet, où l'on trouvera en outre un lexique français-anglais et anglais-français. *Résumé ADBS*.

📖 FOENIX-RIOU, Béatrice. Deux outils pour les référenceurs. *Bases/Netsources*. Novembre/Décembre 2002, n° 41, p. 7.

Présentation des outils de référencement proposés sur Outiref.com (<http://www.outiref.com/>) : audit de page web et de mot-clé. A.M.

☞ FOENIX-RIOU, Béatrice. Outils de recherche : bilan de l'année 2001. *Netsources*. Novembre/Décembre 2001, n° 35, p. 1-5.

L'auteur analyse les bouleversements et tendances qui ont touché le monde des outils de recherche au cours de l'année 2001. Cette année a été marquée par la disparition de plusieurs outils et l'apparition de nouveaux moteurs aux concepts originaux (Kartoo, Wisenut, Teoma...). L'index de Google ne contenait alors que 2 milliards de pages et entreprenait d'indexer des fichiers autres que HTML. Le rafraîchissement rapide des index devenait une priorité tandis que le principe du référencement payant et des emplacements publicitaires se généralisait. A.M.

☞ GARSHOL, Lars Marius. Metadata ? Thesauri ? Taxonomies ? Topic maps ! Making sense of it all. *Journal of Information Science*, 2004, vol. 30, n°4, p. 378-391.

Cet article vise à mieux comprendre l'apport d'outils d'organisation d'information, nouveaux et traditionnels. Il détaille et compare : les métadonnées, les classifications thématiques telles que les vocabulaires contrôlés, les taxonomies, les thésaurus, les classifications à facettes puis les cartes de topiques. Il ressort que les cartes de topiques vont au-delà des solutions traditionnelles car elles fournissent un cadre dans lequel ces dernières peuvent être représentées comme elles sont, mais aussi étendues de façon à améliorer la recherche d'information. *Bulletin bibliographique INTD*, vol. 28, n° 10, Décembre 2004, notice n°022951, p. 44.

✍ IVANCIUC DENIAU, Alina. *Moteurs de recherche et restitution de l'information dans les grandes entreprises*. Mémoire présenté en vue d'obtenir le DESS en sciences de l'information et de la documentation spécialisées, Paris, INTD, 2003, 154 p.

Après avoir replacé les moteurs de recherche dans le contexte de la recherche d'information et les langages documentaires (notamment le thésaurus), compte tenu des mutations de ces dernières années (taxonomies, ontologies, Topic Maps), la première partie du mémoire se propose de décrire le fonctionnement de ces outils issus de la recherche en traitement automatique du langage (TALN). La définition du TALN, en soulignant les apports de chaque discipline impliquée avec un éclairage particulier sur la linguistique, est suivie d'une typologie des produits présents dans les grandes entreprises. Le fonctionnement des moteurs de recherche est décrit ensuite à travers les opérations effectuées par les moteurs de recherche linguistiques pour traiter la masse d'information textuelle lors de l'indexation. Cette description prend pour exemple un produit particulier : « K2 Enterprise » de la société Verity. La seconde partie retrace l'audit effectué afin de mettre en place des améliorations dans un portail d'entreprise dont la recherche est gérée par le moteur de recherche Verity K2 : la Cyberthèque de la Direction des Systèmes d'Information de la branche Banque de Détail de la Société Générale, portail de veille technologique et concurrentielle. *Résumé d'auteur*.

☞ JACQUET, Christophe. Métadonnées et Dublin Core. *Openwebgroup*, [en ligne], 2003, [référence du 8 juin 2005]. http://www.openweb.eu.org/articles/dublin_core/.

Les métadonnées du Dublin Core permettent de donner des informations à propos de pages Web. Dans cet article tiré d'un blog sur les standards du web, l'auteur décrit la forme de ces descriptions, ainsi que la façon de les intégrer dans des pages Web. A.M.

☞ KEISER, Barbie. Enhancing the value of your web site. *Online Magazine*. January/February 2002, vol. 26, n° 1, p. 54-60.

L'auteur, consultante indépendante, propose des astuces pour améliorer la valeur d'un site internet. Dans cette optique, l'article est ponctué de liens vers de nombreux sites et utilitaires. A.M.

🖨 LAZERGES, Alexandre. Amazon, histoire d'un géant. *JDNet Solutions*, [en ligne], 2000, [référence du 14 août 2005]. <http://www.journaldunet.com/0008/000825amazon2.shtml>

L'auteur retrace l'histoire d'Amazon depuis sa création dans un garage jusqu'à son introduction en bourse. A.M.

🖨 LAZERGES, Alexandre. Amazon, le point de vue des éditeurs. *JDNet Solutions*, [en ligne], 29 août 2000, [référence du 14 août 2005]. <http://www.journaldunet.com/0008/000829amazon4.shtml>

L'arrivée d'Amazon dans l'hexagone n'effraie pas outre mesure les éditeurs français. A.M.

🖨 LEVEQUE, Emilie. Les bénéfices d'Amazon plombés par les coûts d'exploitation. *JDNet Solutions*, [en ligne], 28 avril 2005, [référence du 14 août 2005]. <http://www.journaldunet.com/0504/050428amazon.shtml>

Le cybermarchand a publié pour le premier trimestre 2005 un bénéfice net en recul de 30 % à 78 millions de dollars, malgré une croissance de ses revenus de 24 %. En cause : les coûts de livraison. *Résumé d'auteur*.

🖨 MATUCHET, Laurence. Alapage, BOL, La Fnac et Chapitre réagissent à l'arrivée d'Amazon. *JDNet Solutions*, [en ligne], 2000, [référence du 14 août 2005]. <http://www.journaldunet.com/0008/000828amazon3.shtml>

Le secteur de la cyberlibrairie attend de pied ferme l'ouverture d'Amazon.fr. Le Journal du Net est allé prendre le pouls chez Alapage, BOL, la Fnac et Chapitre.com à quelques jours de l'ouverture en France du géant américain. *Résumé d'auteur*.

🖨 MESTRE DE LAROQUE, Amaury. Ces tricheurs qui polluent Google. *L'ordinateur individuel*, [en ligne], 17 février 2005, [référence du 8 juin 2005]. <http://www.01net.com/article/267448.html>

Pour tromper les moteurs de recherche et apparaître en premier dans leurs pages de résultats, certains sites ont recours à des techniques malhonnêtes. Description de ces techniques, de la GoogleBombing et des méthodes développées par les moteurs pour les contrecarrer. A.M.

📖 NEET, H.E.A. La recherche du mot clé : analyse documentaire et indexation alphabétique, Genève : Ed. IES, 1989.

Ce manuel, adressé en priorité aux étudiants en sciences de l'information, présente en dix-huit chapitres les développements de l'indexation. A.M.

🖨 RICHARD, Raphaël. Référencement : le palmarès des sites de campagne. *Le Journal du Net*, [en ligne], 19 avril 2002, (référence du 8 juin 2005].
<http://www.journaldunet.com/dossiers/elections/020419referencement.shtml>

Etude réalisée par l'agence de Webmarketing CVFM sur le référencement dans les outils de recherche des sites des candidats à l'élection présidentielle de 2002. Les pages dynamiques obtiennent les meilleurs résultats. A.M.

📖 RIPERT, Benjamin. *Métier : Webmaster, kit de survie*, Paris : Dunod / 01 Informatique, 2004, 238 p.

Ce manuel s'adresse principalement aux webmasters à qui il propose une synthèse claire et structurée des principaux aspects du métier. Il offre un panorama de solutions techniques opérationnelles et au-delà, s'intéresse également au travail en équipe et à la gestion de projet. Le huitième chapitre « Attirer et fidéliser les internautes » expose les aspects techniques qui concernent le professionnel : les méthodes de référencement (p. 228-231), la gestion d'une newsletter et l'analyse des statistiques de fréquentation. A.M.

📖 ROUMIEUX, Olivier. Référencement et positionnement : gérer la facture numérique. *Archimag*. Avril 2004, n° 153, p. 30-31.

De la soumission au positionnement de son site web, les offres payantes deviennent incontournables. Ce qui ne veut pas dire qu'elles remplacent forcément les techniques gratuites. Petite revue du champ des possibles. *Résumé d'auteur*.

📖 SALAÛN, Jean-Michel. *Marketing des bibliothèques et des centres de documentation*. Paris : Editions du Cercle de la Librairie, 1992, 132 p. 📖

Conçu comme un manuel pratique, cet ouvrage se veut avant tout « une initiation aux techniques de marketing adaptées aux bibliothèques et aux centres de documentation » (cf. Avertissement). Il s'adresse essentiellement aux professionnels des bibliothèques et de la documentation pour leur fournir une première approche critique du marketing. Il s'intéresse successivement à la définition du marketing, à l'analyse de l'activité et de l'environnement des services d'information, aux outils adaptés aux missions et stratégies des centres de documentation et enfin au plan marketing. A.M.

🖨 SULLIVAN, Danny. Death of a meta tag. *Searchenginewatch*, [en ligne], 2002, [référence du 24 août 2005]. <http://searchenginewatch.com/sereport/article.php/2165061>

Cet article retrace l'histoire de la meta Keywords depuis son apparition (Infoseek et AltaVista furent les premiers moteurs à en tenir compte), jusqu'à sa chute, provoquée par les abus du spamdexing. A.M.

🖨 UNIVERSITE DE MONTREAL. Terminologie de base en sciences de l'information : volets 1 et 2. Ecole de bibliothéconomie et des sciences de l'information (EBSI), [en ligne], 16 octobre 2002, [référence du 6 septembre 2005].
<http://www.ebsi.umontreal.ca/termino/index.htm>

Terminologie de l'information-documentation mise en ligne sur le serveur de l'EBSI. A.M.

WEBOGRAPHIE

Les sites sont classés par grands thèmes puis, au sein de chacun d'eux, par ordre alphabétique. Le drapeau qui suit chaque lien indique la langue de rédaction du site.

○ Glossaires sur les moteurs de recherche et le référencement

Comment ça marche. [référence du 08 juin 2005], <http://www.commentcamarche.net/> ■ ■

Cette célèbre encyclopédie libre est une mine d'informations pour tout ce qui touche à l'informatique. On y trouve des définitions sur le référencement et le *web marketing*. A.M.

Dico Du Net. [référence du 18 août 2005], <http://www.dicodunet.com/> ■ ■

Ce dictionnaire collaboratif sur les technologies électroniques permet des recherches par mot-clé, par abécédaire ou par thème. Le site propose un blog et des flux XML au format RSS contenant les dernières définitions ou dernières actualités ajoutées. Voir en particulier les actualités du référencement : <http://www.dicodunet.com/actualites/referencement/> et les définitions de termes liés au référencement : <http://www.dicodunet.com/definitions/referencement>. A.M.

Sumhit. [référence du 18 août 2005], <http://www.sumhit-referencement.com/savoir-lexique.asp> ■ ■

Lexique du référencement et des moteurs de recherche. A.M.

○ Marketing sur Internet

Abc-netmarketing. [référence du 17 septembre 2005], <http://www.abc-netmarketing.com> ■ ■

Fondé par Bertrand Bathelot, ce site s'intéresse essentiellement aux problématiques et techniques du marketing interactif. Il regroupe actuellement une communauté de plus de 10 000 professionnels. Outre des articles, des propositions de formation et le suivi des actualités, il se révèle particulièrement intéressant pour ses glossaires spécialisés sur le marketing traditionnel, le marketing interactif et le référencement. A.M.

○ Métadonnées

Dublin Core. [référence du 19 juin 2005], <http://dublincore.org/> ■ ■

Site officiel du *Dublin Core Metadata Initiative*. A.M.

Educnet. Dossier « Métadonnées, normes et standards ». [référence du 17 septembre 2005] <http://www.educnet.education.fr/dossier/metadata/default.htm> ■ ■

Les objectifs sont clairement exprimés. Il s'agit de montrer l'intérêt des métadonnées et les travaux en cours, d'orienter l'internaute vers les sites spécialisés sur ce sujet et de lui donner des points de repère pour utiliser le Dublin Core. Une bibliographie et un glossaire étayent le dossier. A.M.

Initiative de métadonnées du Dublin Core. [référence du 19 juin 2005], <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm> ■ ■

Traduction française du guide d'utilisation du Dublin Core. A.M.

W3C [référence du 19 août 2005], <http://www.w3.org/> ■ ■

Site officiel du World Wide Web Consortium (W3C), organisation chargée de développer des technologies interopérables sur le web. Le site se veut un carrefour d'échanges et de communication. Il propose entre autres des actualités et des rapports techniques sur des sujets aussi divers que le HTML, les métadonnées, le XML, le web sémantique. A.M.

○ Sites spécialisés sur les outils de recherche et le référencement

Abondance.com. [référence du 10 août 2005], <http://www.abondance.com/> ■ ■

Abondance est LE site de référence francophone sur l'actualité des moteurs de recherche, le référencement et la promotion de sites web. Il est administré par Olivier Andrieu, consultant Internet.

On y trouve entre autres des dossiers, des articles, des études, des méthodologies, un forum, et une suggestion de liens. Abondance propose également une lettre d'information gratuite et hebdomadaire : *Actu-moteurs*, ainsi qu'une lettre d'information payante mensuelle : *Recherche et référencement*. Les mises à jour sont fréquentes. A.M.

Annuaire du référencement. [référence du 08 juin 2005], <http://www.annuaire-referencement.com/> ■ ■

La société Aldeis, qui développe des produits dédiés aux webmasters pour optimiser la visibilité des sites sur Internet, a créé un annuaire des sites consacrés au référencement. A.M.

Asktibbs. [référence du 17 septembre 2005], <http://www.asktibbs.com/> ■ ■

Ce site se présente comme un guide de recherche. Il comprend un annuaire et des fiches techniques sur les outils de recherche, une FAQ et des dossiers pratiques sur la recherche d'information, des entretiens avec des spécialistes du domaine, une lettre d'information consacrée à l'Intelligence Economique, à la veille et au Knowledge Management. Enfin, les professionnels de l'information/documentation ont la possibilité de déposer leur CV dans l'espace professionnel. A.M.

Baromètre référencement [référence du 17 septembre 2005]. <http://www.barometre-referencement.com> ■ ■

Ce baromètre Weborama/Takezo/Brioude repose sur les données recueillies auprès de 2 millions d'internautes par jour. Il dévoile l'évolution des moteurs sur les douze derniers mois. A.M.

IDF.net. [référence du 26 juin 2005], <http://www.idf.net/mdr> ■ ■

Ce site commercial fournit des informations techniques sur les moteurs de recherche francophones. Le glossaire spécialisé n'est plus mis à jour depuis 2000. A.M.

Indicateur. [référence du 17 septembre 2005], <http://www.indicateur.com/> ■ ■

Indicateur.com est un site d'information sur les outils de recherche. Il se compose de trois parties : l'actualité (revue de presse et articles), les dossiers spécialisés (sur le référencement et le positionnement, les moteurs de recherche, les chiffres et classements, la veille et la recherche d'information) et enfin les guides (annuaires thématiques et géographiques). Le site est aussi à l'origine de la création de Google'World, un annuaire spécialisé sur Google. A.M.

Moteurzine. [référence du 19 août 2005], <http://www.moteurzine.com> ■ ■

Lettre d'information de Enfin.com, Moteurzine se définit comme " le bulletin francophone des spécialistes des outils de recherche ". Dans cette newsletter éditée tous les 15 jours, des chroniqueurs issus du monde professionnel partagent leurs réflexions sur les outils de recherche francophones et internationaux et le référencement. A.M.

NetMechanic. [référence du 20 août 2005], <http://netmechanic.com/> ■ ■

Fondée en 1996 par une équipe de professionnels du web, cette société américaine est un fournisseur indépendant d'outils d'évaluation, d'optimisation, de maintenance et de promotion de sites web. Les utilitaires sont proposés à l'essai ou à l'achat. La souscription à la newsletter est elle, en revanche, totalement gratuite. A.M.

PromoWeb : le guide du référencement. [référence du 17 septembre 2005],

<http://www.promo-web.org> ■ ■

Ce site apporte des réponses aux questions les plus fréquentes sur la promotion des sites web, les stratégies de référencement, l'optimisation des pages, l'inscription du site dans les annuaires et moteurs de recherche et enfin le suivi du référencement. A.M.

Référencement, Design et Cie. [référence du 17 septembre 2005],

<http://s.billard.free.fr/referencement/> ■ ■

Ce blog administré par Sébastien Billard, conseiller référencement chez Sumhit, traite des problématiques inhérentes au référencement, à la recherche d'information, à l'accessibilité et au design des sites web ainsi qu'au marketing. A.M.

Revue du référencement. [référence du 17 septembre 2005], [http://www.revue-](http://www.revue-referencement.com/)

[referencement.com/](http://www.revue-referencement.com/) ■ ■

Destinée aux professionnels, la revue du référencement est une publication mensuelle gratuite, axée sur les politiques de référencement. Elle propose des dossiers, études, interviews, chiffres-clés sur les principaux outils de recherche français. Elle est éditée par la société @position, spécialisée dans le positionnement et la visibilité des sites sur Internet. A.M.

Sam-Mag. [référence du 17 septembre 2005], <http://www.sam-mag.com/default.aspx?> ■ ■

Edité par Referenceur.com, ce site est dédié aux webmasters et à la net-économie. Il contient, entre autres, des dossiers et des articles reprenant des avis et expériences de professionnels. Il est possible de s'abonner à la lettre de diffusion. A.M.

Search Engine Showdown. [reference du 17 septembre 2005], <http://www.searchengineshowdown.com> 

Ce site anglophone est développé par Greg R. Notess, bibliothécaire de formation, consultant et contributeur régulier de la revue *Online*. Il offre des ressources spécialisées sur les outils de recherche : données comparatives sur les fonctionnalités des principaux outils de recherche, analyses sur la taille des bases de données des moteurs de recherche, la fréquence de leur "rafraîchissement" et les incohérences de leur fonctionnement. On y trouve aussi des fiches techniques sur des outils de recherche. A.M.

Search Engine Watch. [référence du 17 septembre 2005], <http://searchenginewatch.com> 

Destiné en priorité aux professionnels (experts, webmestres et référenceurs), ce site anglophone est maintenu par Danny Sullivan et Chris Sherman, spécialistes de l'utilisation et de l'évolution du marché des outils de recherche et de l'assistance au référencement. Le site recense les ressources disponibles sur le Web concernant les moteurs de recherche. Il est également possible de s'abonner à une lettre mensuelle (*Search engine report*) et à une lettre d'information quotidienne (*Searchday*), toutes deux gratuites. A.M.

Secrets2moteurs. [référence du 17 septembre 2005], <http://www.secrets2moteurs.com> 

Secret2moteurs permet de suivre les principales tendances et évolutions d'audience et de trafic sur les moteurs de recherche. A.M.

Web Rank Info. [référence du 08 juin 2005], <http://www.webrankinfo.com/> 

Ce site indépendant d'information et de conseils en référencement est édité par Olivier Duffez, consultant en référencement. Il est spécialisé dans les techniques d'optimisation du référencement sur Google, Yahoo Search et MSN Search. Outre des conseils et des actualités, il propose des articles et études sur le référencement, des outils, un forum, des FAQ, un lexique ainsi qu'un annuaire. On regrettera simplement l'absence d'indication de dernière mise à jour. A.M.

○ Sites de référenceurs professionnels

L'offre en matière de référencement est pléthorique. Ne sont indiqués ici, que les sites professionnels mis à contribution dans le cadre de ce mémoire.

Adcom. [référence du 17 septembre 2005], <http://www.adcom.fr> 

Fondé en 1997, cette société d'audit de conseil et de stratégie présente ses prestations. Une rubrique « expertise » regroupe des fiches synthétiques sur les problématiques du référencement, une autre, intitulée « Outils », propose des sites, des applicatifs et scripts sur le sujet. A.M.

@position. [référence du 17 septembre 2005], <http://www.aposition.fr> 

Cette société est spécialisée dans le positionnement et l'optimisation de la visibilité des sites sur Internet. Elle édite « La Revue du référencement ». A.M.

Brioude Internet Référencement. [référence du 17 septembre 2005], <http://www.referencement-2000.com> ■ ■

Prestataire spécialisé en référencement, positionnement publicitaire et achat de mots clés, suivi de positionnement et de trafic, Brioude participe activement à la création d'applicatifs destinés aux référenceurs comme Outiref ou Visiref. *A.M.*

Referenceur. [référence du 17 septembre 2005], <http://www.referenceur.com/> ■ ■

Referenceur.com, qualifié ISO 9001, propose des solutions de référencement professionnel. Il est possible d'y suivre les actualités du référencement. *A.M.*

Webformance. [référence du 17 septembre 2005], <http://www.webformance.fr> ■ ■

Créé en France en 1996, Webformance préside depuis 2003 la *Search Marketing Association*, l'association des professionnels européens. *A.M.*

GLOSSAIRE

Achat de mots-clés : offre commerciale proposée par les outils de recherche aux annonceurs. L'achat de mots-clés, généralement via un système d'enchères, leur permet de se positionner sur les premières pages de résultats à la suite d'une requête.

Adresse IP : équivalent numérique du nom de domaine.

AdWords : c'est le nom donné au programme de positionnement payant de Google. Il est basé sur l'achat de mots-clés aux enchères.

Agent : voir robot.

Algorithme : suite d'opérations composant un programme en vue d'exécuter une tâche spécifique dans un nombre défini d'étapes.

Algorithme de pertinence : routine de programme qu'utilisent les outils recherche pour déterminer la pertinence d'un site ou d'une page Web en fonction de la requête saisie par un internaute. Chaque outil de recherche se rapporte à un algorithme différent.

Annonces : société ou organisme qui lance une opération de communication dans l'objectif de promouvoir ses produits/services et/ou sa marque.

Annuaire (directory) : outil de recherche reposant sur une base de données de fiches descriptives (URL et description) de sites web. Ces-derniers sont classés dans des rubriques (ou catégories) selon une organisation hiérarchique.

Araignée (spider) : voir robot.

Balise (tag) : marque spécifique introduite en début et fin de commande HTML, identifiée par les signes inférieur (<) et supérieur (>) signalant au programme qui interprétera le code qu'une commande spécifique doit être exécutée.

Balise meta (meta tag) : balise HTML permettant d'intégrer des données uniquement destinées aux robots des moteurs de recherche dans des pages Web. Les principales balises méta sont les balises « title », « description » et « keywords ».

Balise meta Description : balise meta donnant une description d'une page Web. La description apparaît souvent dans les résultats des moteurs de recherche en tant que résumé.

Balise meta Keywords : balise meta définissant les mots-clés sur lesquels réagiront les moteurs de recherche en réponse à une requête.

Balise Title : balise meta définissant le titre d'une page Web. Il s'agit du texte qui s'affiche dans le navigateur.

Bannière : lien visuel publicitaire apparaissant sur une page web.

Bruit : ensemble de réponses non pertinentes par rapport à la demande.

Cadre / jeu de cadres (*frame* et *frameset*) : système qui permet d'afficher plusieurs pages web indépendantes dans une même fenêtre du navigateur. Un jeu de cadres ou *frameset* regroupe plusieurs zones de page.

Cloaking : technique consistant à générer un contenu HTML différent selon qu'il s'agit d'un visiteur ou d'un moteur de recherche. Considérée comme du spamdexing, elle est interdite par les moteurs de recherche.

Clustering : fonctionnalité d'un moteur de recherche qui consiste à limiter le nombre de pages issues d'un même.

Cookie : au cours d'une connexion, élément inscrit par un serveur dans un fichier d'ordinateur. Il se configure en fonction de l'utilisateur connecté.

Coût d'acquisition : coût auquel revient l'acquisition d'un client.

Coût par clic : type de tarification qui permet à un annonceur de financer un support en fonction du nombre de clics générés par l'action de communication impliquée.

Coût pour mille impressions : montant facturé à un annonceur pour 1000 affichages d'un message publicitaire sur un support.

Crawler : mot anglais pour chenille. Voir robot.

Descripteur : terme ou expression tirés d'un langage contrôlé pour représenter sans ambiguïté, pendant l'indexation, un concept contenu dans un document ou une recherche documentaire. Il est à différencier du mot-clé, extrait du langage libre.

Dublin Core : ensemble normalisé de métadonnées employées pour décrire les ressources informationnelles sur Internet.

Fichier inverse : type d'index dans lequel l'accès se fait sur les caractéristiques de description.

Forum : service sur Internet permettant l'échange et la discussion entre internautes sur un thème particulier.

Frame / *frameset* : voir cadre/jeu de cadres.

Google Dance : cette expression désigne la période du mois durant laquelle Google met à jour les valeurs de PageRank© des pages indexées dans sa base.

Head : zone d'en-tête d'une page web contenant des informations sur cette dernière (balises meta et « title »).

HTML (*Hypertext Markup Language*) : langage de description employé pour la rédaction des pages Web et commun à toutes les plates-formes (c'est-à-dire tous les types d'ordinateurs quel qu'en soit le système d'exploitation). Les pages HTML sont composées de balises permettant de distinguer la zone d'en-tête du corps de la page.

Hit : un hit correspond à un seul échange HTTP, c'est-à-dire à un fichier transféré du serveur au navigateur. Il peut s'agir d'une page HTML, d'une image, d'un fichier à télécharger ou d'un message d'erreur.

HTTP (*Hypertext Transfer Protocol*) : le but du protocole HTTP est de permettre un transfert de fichiers localisés grâce à une chaîne de caractère appelée URL entre un navigateur et un serveur web.

Index (index) : d'un point de vue informatique, l'index est une base de données répertoriant toutes les pages web d'un outil de recherche et dans laquelle il puise en fonction des requêtes des internautes pour répondre au mieux à leurs questions.

D'un point de vue documentaire, il définit une liste structurée de termes, le plus souvent alphabétique, comprenant des noms propres, des lieux et des sujets traités dans un document. Les termes sont accompagnés d'une référence permettant de les localiser.

Dans les deux cas, sa finalité est de faciliter la recherche d'information.

Indexation : selon la définition de l'AFNOR [AFNOR, 2000, p. 402, (08)], « *processus destiné à représenter par les éléments d'un langage documentaire ou naturel des données résultant de l'analyse du contenu d'un document ou d'une question. On désigne également ainsi le résultat de cette opération.* ».

Ce terme définit également l'action du robot d'un moteur de recherche qui consiste à lire des documents sur le web et à les analyser (syntaxe, liens, etc.) pour les ajouter à sa base de données (son "index").

Indexation automatique (automatic indexing) : indexation accomplie par des moyens informatiques.

Indice de densité d'un mot-clé : il équivaut au nombre de fois où un mot-clé apparaît dans un texte, divisé par le nombre de mots total du texte. On parle aussi d'occurrence.

Indice de popularité : indice proportionnel au nombre de liens pointant vers une page ou un site web : plus il y en a, plus le site est réputé « populaire » et par conséquent pertinent par les moteurs de recherche.

Langage documentaire : ensemble structuré de termes normalisés employé pour refléter le contenu des documents. Un langage documentaire est dit « contrôlé » lorsque ces termes doivent obligatoirement être utilisés à l'exclusion des autres lors de l'indexation ou de la recherche.

On distingue deux catégories de langages documentaires : les langages à structure hiérarchique, (classifications) et les langages analytiques ou à structure combinatoire (thésaurus).

Lien : passerelle entre deux éléments multimédias. Selon les conventions Internet, un lien est bleu et souligné.

Lien contextuel : lien publicitaire et ciblé en fonction de la thématique d'une page de contenu d'un site web.

Lien sponsorisé : lien publicitaire textuel profitant d'un emplacement privilégié (avant les résultats naturels) sur les outils de recherche suite à une requête. L'achat des mots-clés s'effectue généralement selon un système d'enchères et est basé sur un mode de facturation au coût par clic ou au coût par mille.

Fichier log (log file) : fichier journal enregistrant les événements se produisant dans un système.

Métadonnées (*metadata*) : ce sont des données sur les données. Elles fournissent des informations sur des ressources électroniques. Elles permettent de gérer les ressources décrites (suivi du cycle de vie : création, modification, archivage), d'en faciliter l'accès, de suivre l'utilisation et le respect des droits et des conditions qui y sont associées. On distingue actuellement plusieurs standards de métadonnées dont le RDF (Resource Description Framework), le TEI (Text Encoding Initiative), les syntaxes « meta » HTML et Dublin Core, le DTD EAD (Encoding Archival Description), etc.

Méta-moteur : logiciel ou site web qui interroge simultanément plusieurs outils de recherche et en synthétise les résultats.

Mot-clé (*Keyword*) : d'un point de vue documentaire, terme servant à caractériser le contenu sémantique d'un document, pour en permettre la localisation et l'accès. Il ne doit pas être confoncu avec un descripteur, qui est un terme normalisé dans un thésaurus.
D'un point de vue informatique, un mot-clé est un terme saisi par l'internaute pour préciser sa requête sur un outil de recherche.

Moteur de recherche (*search engine*) : outil de recherche de documents sur le Web. Il recense de façon automatique des pages Web via des robots. Un module d'indexation construit automatiquement l'index des termes à partir des résultats tirés de l'étape précédente ; l'interrogation suppose un module de recherche spécifique sur les index ainsi créés.

Navigateur (*browser*) : logiciel employé pour naviguer sur les réseaux informatiques, notamment l'Internet et le Web.

Nom de domaine : traduction d'une adresse IP (composée de différents chiffres) en chaîne de caractères facilitant la mémorisation et l'identification.

Norme : ensemble de règles de conformité édictées par un organisme de normalisation institutionnelle, comme l'AFNOR (Association Française de Normalisation) au niveau français.

Opérateur booléen : opérateur logique (ET/OU/SAUF, opérateurs de proximité ou adjacence, opérateurs numériques) permettant de lier entre eux des mots-clés selon une logique mathématique.

Optimisation : moyen consistant à adapter les pages d'un site web aux critères d'indexation des moteurs de recherche dans le but d'améliorer son référencement naturel.

Outil de recherche (*retrieval tool*) : instrument de recherche utilisé sur Internet. Il existe trois grandes familles d'outils de recherche : les annuaires, les moteurs de recherche et les métamoteurs.

Page alias : page générée pour optimiser le référencement d'un site web sur un ou plusieurs mots-clés et/ou un moteur donné. La création de pages alias ou fantômes est mal perçue par les moteurs de recherche.

Page dynamique : une page dynamique affiche des données différentes en fonction du contexte, en puisant dans une base de données. On la reconnaît à son URL comprenant des variables (index.php?x=1&y=2), bien que la technique de l'*URL rewriting* permette d'y suppléer.

PageRank® : indice de popularité d'une page web créé par Google. Le PR est calculé selon un algorithme très recherché : il juge la pertinence d'une page en fonction du nombre et de la qualité des liens pointant vers elle.

Page satellite : page de pré-entrée sur un site, optimisée spécifiquement pour un moteur de recherche.

Page vue : page web totalement téléchargée sur le poste de l'internaute.

Positionnement (*positionning*) : moyen de placer un site ou une page web dans les premières positions des résultats d'une requête sur les moteurs de recherche.

Push : technique d'envoi d'information en fonction de critères déterminés par l'internaute.

Référencement (*referencing*) : ensemble d'opérations (insertion de métadonnées, enregistrement dans les outils de recherche, veille) visant à favoriser l'accessibilité d'un site web et à améliorer sa visibilité dans les résultats des outils de recherche pour une liste de mots-clés donnée.

Référencement automatique ou naturel : méthode d'indexation d'un site web ayant recours à l'utilisation de logiciels, lesquels permettent un référencement rapide et automatique d'un site dans la base de données de plusieurs moteurs de recherche. Cette méthode traite l'aspect quantitatif du référencement.

Requête ou recherche : mot ou expression saisis par un internaute sur un outil de recherche pour trouver une réponse à sa question sur des pages Internet.

RSS (*Really Simple Syndication* ou *Rich Site Summary*) : moyen d'extraire d'un site web un contenu régulièrement mis à jour. Un fichier RSS est un simple fichier texte au format XML comportant la description synthétique du contenu. Il y a deux raisons d'utiliser RSS : pour récupérer les nouvelles d'un site d'information proposant un flux RSS ou bien afficher de manière dynamique un résumé des dernières news de ce fil sur un autre site.

Retour sur investissement : rapport entre le coût d'acquisition des clients (budget alloué à une campagne) et le chiffre d'affaires réalisé grâce à cette action.

Robot de recherche (*search engine*) : module de collecte automatique utilisé par les moteurs de recherche pour rassembler des données.

Search marketing : terme anglais qui signifie « marketing des outils de recherche ». Il regroupe les techniques classiques d'optimisation et de capitalisation de trafic généré par les outils de recherche (moteurs et annuaires) dont le référencement naturel ou payant.

Silence : définit, dans les résultats d'une recherche, l'absence d'informations et de documents malgré leur pertinence.

Spamdexing : ensemble des techniques et méthodes abusives de référencement (soumissions multiples, répétitions de mots-clés, pages satellites, duplication de pages, *cloaking*, etc.) consistant à tromper les outils de recherche dans le but d'optimiser le positionnement d'un site parmi les résultats de recherche.

Soumission : fait de proposer, via un formulaire, l'inscription d'un site web à un annuaire de recherche. Avant d'être enregistré dans l'index, le site doit être validé par un netsurfeur.

Soumission payante : inscription d'un site web dans un annuaire de recherche via un formulaire d'inscription payant contre la garantie du délai de traitement.

Spider : araignée en français. Voir agent ou araignée.

Standard : ensemble de préconisations développées par un groupe représentatif d'utilisateurs.

Syndication : technique permettant d'afficher automatiquement dans un site des données issues d'autres sites. Son intérêt est de suivre des pages et sites qui évoluent régulièrement.

Tag : voir balise.

Taux de transformation (taux de vente ou d'inscription) : pourcentage d'internautes ayant effectué un achat ou une inscription sur un site web, par rapport au nombre total de visites sur le site. Un taux de transformation de 5 % signifie que sur cent visiteurs, cinq ont finalement réalisé un achat.

Texte intégral (full text) : technique d'indexation informatique dans laquelle tous les mots d'un texte, à l'exception éventuelle des mots vides, font l'objet d'une entrée dans le fichier index.

Toile : mot français désignant le World Wide Web ou W3.

Trafic : il équivaut au nombre total de visites reçues par un site Internet sur une période déterminée.

Trafic qualifié : se dit des visiteurs intéressés par l'activité et les ventes d'un annonceur.

URL (Uniform Resource Locator) : adresse d'une ressource électronique sur un ordinateur local ou sur le Web.

URL exotique : une URL est dite exotique lorsqu'elle contient dans son intitulé des caractères spécifiques comme le point d'interrogation (?) ou l'esperluette (&) et une à plusieurs variables. Elle est typique des pages dynamiques.

Webmaster : voir webmestre.

Webmestre (webmaster) : responsable technique d'un site web.

W3C (World Wide Web Consortium) : créé en octobre 1994 par Tim Berners Lee, l'inventeur du Web, ce consortium constitué de centaine de membres (grandes sociétés de l'informatique et des télécoms, éditeurs, instituts de recherche) a pour mission de guider les évolutions technologiques du réseau mondial grâce à l'élaboration de spécifications techniques et de réfléchir à ses développements futurs.

XML (eXtensible Markup Language) : langage de balisage extensible défini par le W3C pour la création de formats de documents courants. Les données XML séparent nettement le contenu, le style et la structure des documents ce qui facilite la présentation des données dans différents formats et applications. XML est un métalangage, c'est-à-dire un ensemble de règles permettant de définir d'autres langages balisés spécialisés. XHTML ou RSS sont des langages basés sur XML.

ANNEXES

Annexe 1 : capture d'écran d'une page produit Alapage.com

The screenshot shows the Alapage.com website interface. At the top, there's a navigation bar with the Alapage logo and a banner for a 2-month free subscription. Below this is a menu with categories like LIVRES, MUSIQUES, DVD/VIDÉO, etc. The main content area features a large banner for 'PRIX IMBATTABLES' and a search bar. The product page for 'REFERENCEMENT SUR LE NET' by Sandrine Saporta is displayed, showing the book cover, price (23.75€), and a 5% discount. The page also includes a 'RECHERCHE RAPIDE' section, a 'SUR CET ARTICLE' section with characteristics and comments, and a 'RECOMMANDATIONS' section. The footer contains a navigation menu, contact information, and a copyright notice for 1996-2004 Alapage.com.

Annexe 2 : en-tête HTML d'une page produit Alapage.com

```
<html>
<head>
<SCRIPT LANGUAGE="JavaScript1.1"><!-- var
random_number="ord="+ (Math.round(Math.random()*10000000000))+ "?"; //-->
</SCRIPT>
    <script language="JavaScript1.2" src="/outils/popunder.js"></script>
    <!--DEBUT HEAD AIMFAR-->
    <script language="javascript">var wp_loaded=0;</script>
    <!--FIXME aimfar.js location -->
    <script language="javascript" src="/inc/pub/aimfar.js"></script>
    <!--FIN HEAD AIMFAR-->
<link rel="SHORTCUT ICON" href="/favicon.ico" />
<title>alapage.com - Livre: REFERENCEMENT SUR LE NET</title>
<meta name="description" content="REFERENCEMENT SUR LE NET">
<meta name="keywords" content="REFERENCEMENT SUR LE NET">
<script language="JavaScript">
<!--
var beginDate = new Date();
function putTime (){
    var finalDate = new Date ()
    var dureeHtml = htmlDate.getTime() - beginDate.getTime();
    var dureeTotale = finalDate.getTime() - beginDate.getTime();
    document.getElementById("test").innerHTML += "html : "+dureeHtml+" gets :
"+dureeTotale;
}
var random_number = "ord=" + (Math.round (Math.random () * 10000000000)) + "?";
var encours  = new Image();
encours.src  = "/images/alapage/habillage/ch_en_cours.gif";
encours.width  = "200";
encours.height = "48";
var fin  = new Image();
fin.src  = "/produits/commun/img/pix_trans.gif";
fin.width  = "100";
fin.height = "100";
function bascule (from, to)
{
    document[from].src = to.src;
}
//-->
</script>
<script language="javascript">var AlapageURL = "http://www.alapage.com";</script><script
language="JavaScript" type="text/javascript" src="/cart/js/cookies.js"></script>
<script language="JavaScript" type="text/javascript" src="/cart/js/configuration.js"></script>
```

```
<script language="JavaScript" type="text/javascript"
src="/cart/js/remote_shoppingcart_func.js"></script>
<!-- idevelay -- Hitech Lot 2 -- ajout pour 123Crédit -->
<script language="JavaScript" type="text/javascript" src="/cart/js/mod_123.js"></script>
<!-- FIN idevelay -- Hitech Lot 2 -- ajout pour 123Crédit -->
<script language="JavaScript" type="text/javascript" src="/mx/templates/popup_layer.js"></script>
<script language="JavaScript">
function changeIframeHeight (iframeId)
{
    var ifr_height = document.getElementById
("shoppingCartIframe").document.body.scrollHeight;
    document.getElementById ("shoppingCartIframe").style.height = ifr_height;
}
</script>
<link type="text/css" rel="stylesheet" href="/css/lot2.css.php?type=1&cod_part=&tp=F" />
</head>
```


Les internautes intéressés par cet article peuvent être intéressés par :

Lien sponsorisé ([En savoir plus](#)) [Votre avis](#)

- [Création de sites Web pour les PME et les TPE](#)
Solution globale de création de sites e-business clés en main : logiciel d'administration du site, hébergement et panier d'achat pour PME et TPE, à des prix mesurés. 15 à 40 EUR hors taxes par mois, sans frais initiaux ni engagement de durée.
[secure.oxatis.com](#)
- [Communication et collaboration : enjeu stratégique](#)
Découvrez comment Microsoft Exchange s'adaptera au mieux à vos spécificités métiers et améliorera la compétitivité de votre entreprise. Un dossier spécial hors-série, "Lotus Notes vs Exchange Server", de 44 pages est disponible gratuitement.
[www.itpro.fr](#)
- [BoutiqueExpert : Votre boutique en ligne en toute autonomie](#)
Avec BoutiqueExpert, Affichez votre business **sur** Internet. Une solution complète pour créer, gérer et développer votre boutique en ligne. Un prix tout compris : boutique, hébergement, support, nom de de domaine. Version d'évaluation gratuite.
[www.boutique-expert.fr](#)

Langue : Français **Éditeur :** Organisation (5 mai 2005)
Collection : ED ORGANISATION
Format : Broché - 190 pages
ISBN : 2708133020
Dimensions (en cm) : 16 x 1 x 24

Table des matières

► [Table des matières complète](#)

Chroniques et points de vue

Présentation de l'éditeur

Choisir les bonnes techniques de référencement permet d'avoir une présence intelligente et rentable sur Internet. Ce livre, très opérationnel, permet de comprendre comment marchent les référencements et quelles sont les techniques à disposition, même les plus récentes comme les liens sponsorisés ou le trustfeed. Il fournit aux responsables marketing les démarches, conseils, trucs et astuces, avantages et inconvénients de chaque méthode pour choisir en connaissance de cause et être ainsi gagnant sur Internet.

Biographie de l'auteur

Sandrine SAPORTA a géré des grands comptes au sein de la cellule promotion des ventes du groupe Publicis. Elle a co-fondé l'agence de communication interactive Ciblo. net. Elle est titulaire d'une maîtrise ès Lettres Modernes, diplômée d'école supérieure de publicité, et de l'International Advertising Association.

► [Liste complète des chroniques et points de vue...](#)

Les internautes ayant acheté des titres de Sandrine Saporta ont également acheté des titres de :

- [Marie Prat](#)
- [Olivier Andrieu](#)
- [Tara Calishain](#)
- [Jakob Nielsen](#)
- [Nicolas Chu](#)

Rechercher des livres semblables à Référencement sur le net par sujet :

Parcourir les **Livres** :

- [Par prix](#) > [De 20 à 45 euros](#) > [Entreprise, Bourse et Droit](#)
- [Thèmes](#) > [Entreprise et Bourse](#) > [Meilleures ventes](#)
- [Thèmes](#) > [Entreprise et Bourse](#) > [Gestion et Administration](#) > [Meilleures ventes](#)
- [Thèmes](#) > [Entreprise et Bourse](#) > [Gestion et Administration](#) > [Gestion](#) > [Meilleures ventes](#)
- [Thèmes](#) > [Entreprise et Bourse](#) > [Divers](#)
- [Thèmes](#) > [Entreprise et Bourse](#) > [Gestion et Administration](#) > [Gestion](#) > [Divers](#)

Rechercher des livres par sujets :

☐ [E-business](#)

☐ [Entreprise](#)

☐ [Gestion](#)

Trouver les livres correspondants

par exemple, chaque livre doit correspondre au sujet 1 ET au sujet 2 ET

- Vous avez lu ce livre et souhaitez [donner votre avis](#).
- [Corriger](#) les erreurs ou ajoutez des informations sur cet article.

Où en sont vos commandes ?

- Suivez l'état d'avancement de vos [commandes récentes](#)
- Consultez et modifiez vos commandes dans "[Votre compte](#)"

Livraison et retours

- Lire nos [tarifs & options de livraison](#)
- Comment nous [retourner un article](#) ?

Besoin d'aide ?

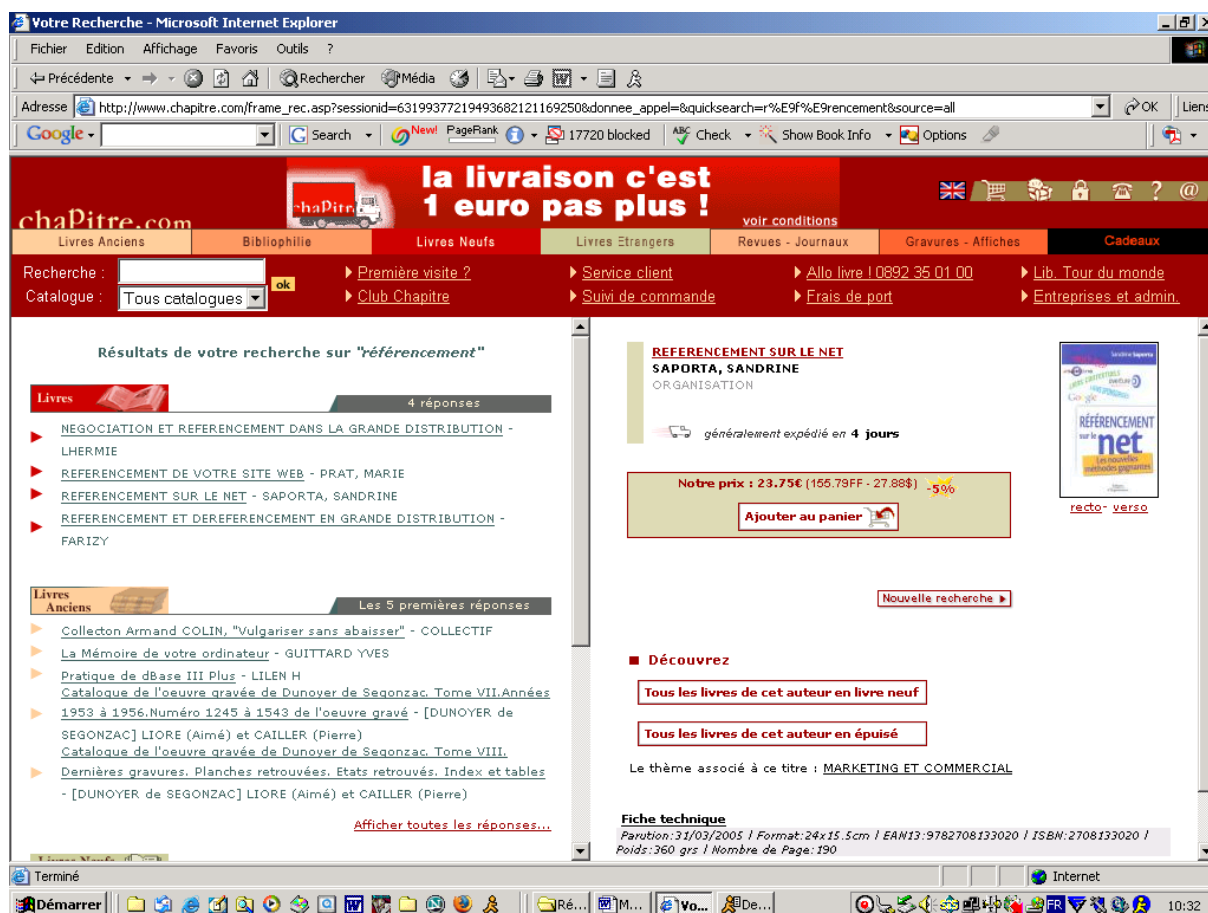
- Vous avez oublié [votre mot de passe](#) ?
- Acheter ou utiliser un [chèque-cadeau](#)
- Consultez nos [pages d'aide](#)

Annexe 4 : en-tête HTML d'une page produit Amazon.fr

```
<html>
<head>
<style type="text/css"><!-- .serif { font-family: times,serif; font-size: small; }
.sans { font-family: verdana,arial,Helvetica,sans-serif; font-size: small; }
.small { font-family: verdana,arial,Helvetica,sans-serif; font-size: x-small; }
.h1 { font-family: verdana,arial,Helvetica,sans-serif; color: #CC6600; font-size: small; }
.h3color { font-family: verdana,arial,Helvetica,sans-serif; color: #CC6600; font-size: x-small; }
.tiny { font-family: verdana,arial,Helvetica,sans-serif; font-size: xx-small; }
.websearch-string { font-family: verdana,arial,Helvetica,sans-serif; font-size: xx-small; }
.listprice { font-family: arial,verdana,sans-serif; text-decoration: line-through; font-size: x-small; }
.price { font-family: verdana,arial,Helvetica,sans-serif; color: #990000; font-size: x-small; }
.tinyprice { font-family: verdana,arial,Helvetica,sans-serif; color: #990000; font-size: xx-small; }
.attention { background-color: #FFFFD5; }
.eyebrow { font-family: verdana,arial,Helvetica,sans-serif; font-size: 10px; font-weight: bold; text-
transform: uppercase; text-decoration: none; color: #FFFFFF; }
A.eyebrow:link { text-decoration: none; }
.breadcrumb { font-family: verdana,arial,Helvetica,sans-serif;font-size: x-small;font-weight:bold; }
.breadcrumb-node { font-family: verdana,arial,Helvetica,sans-serif;font-size: x-small;font-weight:
bold;color: #CC6600; }
.horizontal-search { font-weight: bold; font-size: small; color: #FFFFFF; font-family:
verdana,arial,Helvetica,sans-serif; }
.horizontal-websearch { font-size: xx-small; font-family:verdana,arial,Helvetica,sans-serif;padding-left:
12px; }
.horizontal-advanced-websearch { font-size: xx-small; color: #ffffff; font-
family:verdana,arial,Helvetica,sans-serif; text-decoration: none }
A.horizontal-advanced-websearch: hover { text-decoration: underline }
.topnav { font-family: verdana,arial,Helvetica,sans-serif; font-size: 12px; text-decoration: none; }
.topnav A:link, .topnav A:visited { text-decoration: none; color: #003399; }
.topnav A: hover { text-decoration: none; color: #CC6600; }
.topnav-active A:link, .topnav-active A:visited { font-family: verdana,arial,Helvetica,sans-serif; font-
size: 12px; color: #CC6600; text-decoration: none; }
.tabon a, .tabon a:visited { font-size: 10px; color: #FFCC66; font-family: verdana,arial,Helvetica,sans-
serif; text-decoration: none; text-transform: uppercase; font-weight: bold; line-height: 10px; }
.taboff a, .taboff a:visited { font-size: 10px; color: #000000; font-family: verdana,arial,Helvetica,sans-
serif; text-decoration: none; text-transform: uppercase; font-weight: bold; line-height: 10px; }
.tabon a: hover, .taboff a: hover { text-decoration: underline; }
.tabon div, .taboff div { margin-top: 7px; margin-left: 9px; margin-bottom: 5px; }
.popover-tiny { font-size: xx-small; font-family: verdana,arial,Helvetica,sans-serif; }
.popover-tiny a, .popover-tiny a:visited { text-decoration: none; color: #003399; }
.popover-tiny a: hover { text-decoration: none; color: #CC6600; }
.indent { margin-left: 1em; }
.half { font-size: .5em; }
.list div { margin-bottom: 0.25em; text-decoration: none; }
```

```
.hr-center { margin: 15px; border-top-width: 1px; border-right-width: 1px; border-bottom-width: 1px;
border-left-width: 1px; border-top-style: dotted; border-right-style: none; border-bottom-style: none;
border-left-style: none; border-top-color: #999999; border-right-color: #999999; border-bottom-color:
#999999; border-left-color: #999999; }
.divider { border-top: 1px dashed #cc9; }
.big { font-family: verdana,arial,Helvetica,sans-serif; font-size: ; }
--></style>
<META name="description" content="Référencement sur le net, Sandrine Saporta">
<META name="keywords" content="Référencement sur le net, Livres, Sandrine Saporta, E-business,
Entreprise, Gestion">
<title>Amazon.fr&nbsp;: Livres: Référencement sur le net </title>
</head>
```

Annexe 5 : capture d'écran d'une page produit Chapitre.com



Annexe 6 : code source HTML d'une page produit Chapitre.com

```
<html>
<head>
<title>Votre Recherche</title>
<!-- Begin WEBandSTATS Tag Rev 3.5 part 1-->
<!-- COPYRIGHT AUDIENTIA ALL RIGHTS RESERVED. -->
<script language='javascript'>
<!--
var SITE_ID      = "F6F6FF1F7F";
var PAGE_URL     = escape('http://www.chapitre.com/frame_rec.asp');
var PAGE_NAME    = "ASP-FRAME Moteur Recherche";
var HANDLE_PARAM = "YES";
var PARAMETERS   = "";
var GALLERY_PRODUCTID= "";
var GALLERY_NAME = "";
var scjsexec = "no";
//-->
</script>
<script language='javascript1.1' src='/webandstats_v35.js'></script>
<script language="javascript">
<!--
if (scjsexec=="no") {
var wasa = 'http://apu03c0.audientia.net:80/scripts/stats.asp';
var wasb =
'?SITE_ID=F6F6FF1F7F&REFERRER='+escape(document.referrer)+'&LOCAL_DATE='+
(new Date()).getHours();
wasb +=
'&JS_VERSION=10&PAGE_TITLE='+escape(document.title)+'&PAGE_NAME='+escape(PAGE_N
AME);
wasb +=
'&PAGE_URL='+PAGE_URL+'&HANDLE_PARAM='+HANDLE_PARAM+'&GALLERY_NAM
E='+escape(GALLERY_NAME);
wasb +=
'&GALLERY_PRODUCTID='+escape(GALLERY_PRODUCTID)+'&PARAMETERS='+escape(PA
RAMETERS);
var wasImg = new Image();wasImg.src = wasa+wasb;}};
//-->
</script>
<noscript>

</noscript>
<!-- End WEBandSTATS Tag -->
```

```

<!-- ADHOC -->
<script language="JavaScript">
<!--
window.onerror=dierr;
var tiSiteID = 217;
var tiComplement = "";
document.write('<sc'+ript language=javascript1.1
src="'+((document.location.href.toString().indexOf("/")==5)?"http":"https")+'://go.atraxio.com/decisiv
einsight/trjs.asp"></sc'+ript>');
function dierr()
{return true;}
//-->
</script>
<!-- ADHOC -->
</head>
<frameset rows="135,*" border="0">
    <frame name="haut" scrolling="no" noresize
src="/accueil.asp?page=/menu/m_all.htm&sessionid=5581847443318408311467166&donnee_appel=
">
        <frameset cols="48%,*">
            <frame name="gauche"
src="/asp/panier/search_all.asp?quicksearch=r%E9f%E9rencement&source=ALL&pid=&sessionid=5
581847443318408311467166&donnee_appel=">
                <frame name="droite"
src="/asp/panier/saisie_all.asp?quicksearch=r%E9f%E9rencement&source=ALL&pid=&sessionid=55
81847443318408311467166&donnee_appel=">
            </frameset>
        <noframes>
            <body>
                <p>Cette page utilise des cadres, mais votre navigateur ne les prend pas en charge.</p>
            </body>
        </noframes>
    </frameset>
</html>

```


Annexe 7 : capture d'écran d'une page produit Fnac.com

La plus grande Fnac, disponible en ligne 24h/24

Rechercher : Livres

Référéncement sur le net
Les nouvelles méthodes gagnantes
S. Saporta
guide (broché). Paru en 02/2005
Expédié sous 4 à 8 jours

23,75 € (5% de réduction)

AJOUTER AU PANIER

Les internautes ayant acheté cet article ont également acheté :

- Conduite de projet Web** - Stéphane Bordage - 33,25 €
- Référéncement de votre site Web** - Marie Prat - 20,13 €

En résumé

Mot de l'éditeur
Il est possible de gagner beaucoup d'argent en choisissant les bonnes techniques de référencement. Ce livre, très opérationnel, permet de comprendre comment marchent les référencements et quelles sont les techniques à disposition, qu'elles soient traditionnelles ou nouvelles. Un livre qui fournit aux responsables marketing les démarches, conseils, trucs et astuces, avantages et inconvénients de chaque technique pour avoir une présence intelligente et rentable sur internet.

SOMMAIRE :
INTRODUCTION :
COMMENT LA STRATEGIE "PULL" PERMET D'ETRE GAGNANT SUR INTERNET

I - COMMENT GENERER UN TRAFIC LUCRATIF SUR VOTRE SITE : LES TECHNIQUES DE REFERENCEMENT

1. La différence entre annuaires et moteurs de recherche
2. Les techniques traditionnelles de référencement
3. La soumission / Le référencement par indexation / Le référencement par positionnement

Les nouvelles méthodes de référencement : "Quand payer un clic vous permet de gagner"

Les liens sponsorisés / Les liens contextuels / Le Trust Feed

II - LES LIENS SPONSORISES : LA PREUVE PAR L'EXEMPLE

1. Travailler avec les prestataires
2. Le marché des liens promotionnels

MENTO :
LES ASPECTS JURIDIQUES DU REFERENCEMENT

Fiche détaillée

Auteur	S. Saporta
Editeur	Organisation Eds D'
Date de parution	02/2005
Format	16 cm x 24 cm
ISBN	2708133020

Vos avis

Votre avis nous intéresse : Votre note de 1 à 10 [10 (très bien)]

AJOUTER AU PANIER

Mon panier | **Mon compte** | **Newsletter Fnac** | **Espace Adhérents** | **Aide**

Engagements Fnac.com | L'entreprise Fnac | Recrutement | Fnac dans le monde | Magasins Fnac | Contactez-nous

Conditions générales de vente © Fnac 2005

Annexe 8 : en-tête HTML d'une page produit Fnac.com

```
<script language=javascript>
function DisplayPopUp(theURL)
{
    window.open(theURL,'PopUp','toolbar=no,location=no,menubar=no,status=no,top=0,left=0,
scrollbars=yes,width=530,height=585');
}
</script>
<SCRIPT LANGUAGE=javascript>
<!--
function ShowDTI(url)
{
    window.open(url,'FolderTech','width=820,height=600,left=0,top=0,resizable=yes,toolbar=no,l
ocation=no,directories=no,status=no,scrollbars=yes,menubar=no');
}
//-->
</SCRIPT>
<script language="javascript">
function newWindow(theURL) {
window.open(theURL,'FNAC','toolbar=no,location=no,menubar=no,status=no,scrollbars=yes,width=6
00,height=500');
}
</script>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
<meta http-equiv="PICS-Label" content="(PICS-1.1 "http://www.rsac.org/ratingsv01.html" l gen true
comment "RSACi North America Server" by "webmaster@fnac.com" for "http://www.fnac.com/" on
"1997.06.30T14:21-0500" r (n 0 s 0 v 0 l 0))">
<meta name="MS.LOCALE" content="EN-US">
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<meta name="CATEGORY" content="home page">
<meta name="DESCRIPTION" content="Commandez Référencement sur le net de S. Saporta sur
Fnac.com, votre magasin Fnac 24 heures sur 24, 7 jours sur 7.">
<meta name="KEYWORDS" content="FNAC,Fnac, fnac, livres, livre, LIVRE, bandes dessinées, bd,
BD, bouquin, bouquins,guide, best-sellers, journal, librairie, libraire, roman, écrivain, polar,éducation,
finance, nature, histoire, lettres, éducation, poésies, conte,enfants, informatique, langues, cuisine,
littérature, scolaire, université,concours, prix, art">
<meta name="ROBOTS" content="INDEX,FOLLOW">
<meta name="REVISIT-AFTER" content="2 DAYS">
<!-- LastVisit:16/09/2005 10:38:03 -->
<title>Fnac.com – Livres – Référencement sur le net - S. Saporta</title>
```

```
<link rel="stylesheet" type="text/css" href="/css/fnac.css">
<link rel="stylesheet" type="text/css" href="/css/fnac_base.css">
<link rel="stylesheet" type="text/css" href="/css/fnac_LIV.css">
<link rel="stylesheet" type="text/css" href="/css/fnac_perso.css">
<link rel="stylesheet" type="text/css" href="/css/FDstyle_checkout.css">
</head>
```

Annexe 9 : capture d'écran d'une page produit Dunod.com



Annexe 10 : en-tête HTML d'une page produit Dunod.com

```
<SCRIPT LANGUAGE="JavaScript">
function testEmail(email)
{var exp_email=/.+@.+\./;
return exp_email.test(email);
}</script>
<HTML>
<HEAD>
<TITLE>Dunod/01 Informatique - livre Check-list pour réussir son site web : Les soixante points clés
- Jean-Marc Hardy</TITLE>
<META NAME="description" CONTENT="livre Check-list pour réussir son site web de Jean-Marc
Hardy, Les soixante points clés.">
<META NAME="title" CONTENT="Dunod/01 Informatique - livre Check-list pour réussir son site
web : Les soixante points clés - Jean-Marc Hardy">
<META NAME="keywords" CONTENT="livre bureautique, livres bureautique">
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-1">
<LINK REL="stylesheet" HREF="../dunodstyle.css" TYPE="text/css">
<SCRIPT LANGUAGE="JavaScript">
function MM_swapImgRestore() { var i,x,a=document.MM_sr;
for(i=0;a&&i<a.length&&(x=a[i])&&x.oSrc;i++) x.src=x.oSrc; }
function MM_preloadImages() { var d=document; if(d.images){ if(!d.MM_p) d.MM_p=new Array();
var i,j=d.MM_p.length,a=MM_preloadImages.arguments; for(i=0; i<a.length; i++) if
(a[i].indexOf("#")!=0){ d.MM_p[j]=new Image; d.MM_p[j++].src=a[i];}}
}function MM_findObj(n, d) { var p,i,x; if(!d) d=document;
if((p=n.indexOf("?"))>0&&parent.frames.length) { d=parent.frames[n.substring(p+1)].document;
n=n.substring(0,p); } if(!(x=d[n])&&d.all) x=d.all[n]; for (i=0;!x&&i<d.forms.length;i++)
x=d.forms[i][n];
for(i=0;!x&&d.layers&&i<d.layers.length;i++) x=MM_findObj(n,d.layers[i].document); if(!x &&
document.getElementById) x=document.getElementById(n); return x; }
function MM_swapImage() { var i,j=0,x,a=MM_swapImage.arguments; document.MM_sr=new
Array; for(i=0;i<(a.length-2);i+=3) if ((x=MM_findObj(a[i]))!=null){document.MM_sr[j++]=x;
if(!x.oSrc) x.oSrc=x.src; x.src=a[i+2];}
}function MM_reloadPage(init) { if (init==true) with (navigator) {if
((appName=="Netscape")&&(parseInt(appVersion)==4)) { document.MM_pgW=innerWidth;
document.MM_pgH=innerHeight; onresize=MM_reloadPage; }}
else if (innerWidth!=document.MM_pgW || innerHeight!=document.MM_pgH) location.reload(); }
MM_reloadPage(true);
function MM_showHideLayers() { var i,p,v,obj,args=MM_showHideLayers.arguments;
for (i=0; i<(args.length-2); i+=3) if ((obj=MM_findObj(args[i]))!=null) { v=args[i+2]; if (obj.style) {
obj=obj.style; v=(v=='show')?'visible':(v=='hide')?'hidden':v; } obj.visibility=v; } }
function MM_openBrWindow(theURL,winName,features) {
window.open(theURL,winName,features); }
</SCRIPT>
</HEAD>
```

Annexe 11 : capture d'écran d'une page produit Eyrolles.com

Livre Référencement sur le net - S. Saporta - Les nouvelles méthodes gagnantes - Librairie Eyro - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Recherche Média

Adresse <http://www.eyrolles.com/Accueil/Livre/9782708133020/livre-referencement-sur-le-net.php>

Google Search New! PageRank 17720 blocked Check Show Book Info Options

Eyrolles.com La librairie des professionnels

Panier Aide Site sécurisé Loupe

Accueil Informatique Audiovisuel et Graphisme Sciences et Techniques Entreprise Droit BTP et Construction Loisirs et Vie quotidienne

RECHERCHE Tous les thèmes OK Recherche détaillée

Accueil / Entreprise / Marketing / E-marketing / Référencement sur le net

Fiche

- Caractéristiques
- Extraits du livre
- Présentation par l'éditeur
- Au sommaire
- Avis des lecteurs
- Les internautes qui ont acheté ce livre ont aussi acheté
- Consultez aussi

Votre compte

E-mail :

Mot de passe oublié ? Inscrivez-vous gratuitement !

Abonnements

Pour votre veille bibliographique :

- des fils RSS
- des alertes e-mail

Des services gratuits pour être informé des nouveautés qui vous intéressent... Abonnez-vous !

Aide

Référencement sur le net

Les nouvelles méthodes gagnantes

Sandrine Saporta
Editions d'Organisation
Prix public : 25,00 EUR
Prix eyrolles.com : 23,75 EUR (155,79 FRF)
Réduction : -5%
Expédié dans les 24 heures
En vente dans notre librairie parisienne

Agrandir l'image

Ajouter au panier Ajouter à la présélection

Référencement sur le net - Les nouvelles méthodes gagnantes

- Editeur(s) : Editions d'Organisation - Auteur(s) : S. Saporta - Collection : - Profil : - Niveau :	- Parution : 06/05/2005 - Edition : 1ère édition - Nb de pages : 190 pages - Format : 15,5 x 24 - Couverture : Broché - Poids : 360 g - Intérieur : Noir et Blanc	Type produit : Livre - Langue : Français - ISBN : 2-7081-3302-0 - EAN13 : 9782708133020 - Inclus :
--	---	---

Extraits du livre

- Table des matières** - Fichier PDF (185.9 Ko)
- Introduction** - Fichier PDF (406.7 Ko)
- Partie 1 : Générer un trafic lucratif : les techniques de référencement - Chapitre 2 : Les techniques traditionnelles de référencement - Pages 17 à 35** - Fichier PDF (576.7 Ko)
- Partie 2 : Liens sponsorisés : la preuve par l'exemple - Chapitre 1 : Une agence et ses annonceurs témoignent - Pages 95 à 100** - Fichier PDF (380 Ko)

Conseillez cet ouvrage à un ami

A lire aussi

- Salon : Entrepreneurs
- Interview : Raphaël Goetter
- Interview : Eclipse et JBoss

Dernières parutions de ce thème

L'e-mail marketing

Toutes les nouveautés

- ▶ Trouver un article
- ▶ Frais de port et délais
- ▶ Commander
- ▶ Paiement
- ▶ Sécurité
- ▶ Compte eyrolles.com
- ▶ Suivi de commande
- ▶ Nous contacter
- ▶ Venir à la librairie
- ▶ Sommaire de l'aide

▶ Présentation par l'éditeur

Toutes les techniques pour générer un trafic lucratif sur votre site !

- Un **panorama complet des techniques** de référencement
- Des **fiches pratiques** avec conseils et astuces
- Des **exemples concrets** pour illustrer chaque problématique
- Un **lexique** des termes techniques et un **mémento juridique**

Choisir les bonnes techniques de référencement permet d'avoir une présence intelligente et rentable sur Internet.

Ce livre, très opérationnel, permet de comprendre comment marchent les référencement et quelles sont les techniques à disposition, même les plus récentes comme les liens sponsorisés ou le *trust feed*. Il fournit aux responsables marketing les démarches, conseils, trucs et astuces, avantages et inconvénients de chaque méthode pour choisir en connaissance de cause et **être ainsi gagnant sur Internet**.

▶ Au sommaire

- Avant-propos
- Introduction
- Générer un trafic lucratif : les techniques de référencement
 - La différence entre annuaires et moteurs de recherche
 - Les techniques traditionnelles de référencement
 - Les nouvelles méthodes de référencement : "quand payer un clic vous permet de gagner"
- Liens sponsorisés : la preuve par l'exemple
 - Une agence et ses annonceurs témoignent
 - Le marché des liens promotionnels
 - Mémento juridique
- Glossaire
- Index

▶ Avis des lecteurs

Aucun avis disponible.

Donnez votre avis à propos de ce livre

▶ Les internautes qui ont acheté ce livre ont aussi acheté



La PAO ... comment ça marche?

13,78 EUR

🛒 Ajouter au panier



Les nouvelles formes d'addiction

7,79 EUR

🛒 Ajouter au panier

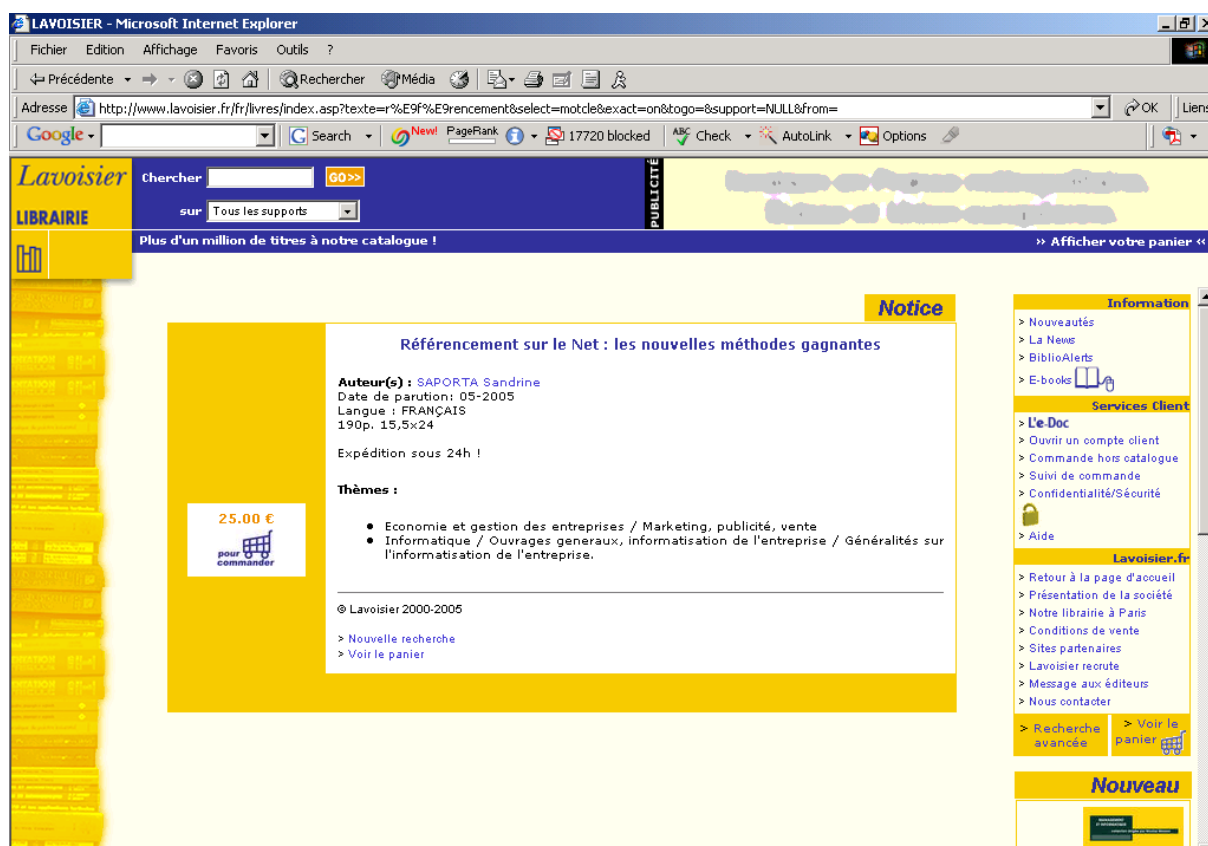
▶ Consultez aussi

- [Entreprise](#) > [Entreprise](#) > [Marketing](#) > [E-marketing](#)
- [Informatique](#) > [Informatique d'entreprise](#) > [Marketing de sites](#)
- [Tous les livres de Sandrine Saporta](#)
- [Salon : Entrepreneurs](#)
- [Interview : Raphaël Goetter](#)
- [Interview : Eclipse et JBoss](#)
- [Informatique](#) > [Informatique d'entreprise](#) > [Intranet et portails](#)
- [Informatique](#) > [Informatique d'entreprise](#) > [Commerce électronique](#)
- [Entreprise](#) > [Entreprise](#) > [Commercial](#) > [E - commerce](#)
- [Entreprise](#) > [Entreprise](#) > [Marketing](#)

Annexe 12 : en-tête HTML d'une page produit Eyrolles.com

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr">
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
  <title>Livre Référence sur le net - S. Saporta - Les nouvelles méthodes gagnantes - Librairie
Eyrolles</title>
  <meta http-equiv="Content-Script-Type" content="text/javascript" />
  <meta http-equiv="Content-Style-Type" content="text/css" />
  <meta http-equiv="Content-Language" content="fr" />
  <link rel="StyleSheet" href="/layout/librairie/css/print.css" type="text/css" media="print" />
  <link rel="alternate stylesheet" type="text/css" media="screen,print"
href="/layout/librairie/css/txtsm.css" title="Texte petit" />
  <link rel="alternate stylesheet" type="text/css" media="screen,print"
href="/layout/librairie/css/txtlg.css" title="Texte grand" />
  <link rel="alternate stylesheet" type="text/css" media="screen,print"
href="/layout/librairie/css/txtlg2.css" title="Texte plus grand" />
  <meta name="keywords" content="Eyrolles, librairie, livres d informatique, nouvelles technologies,
e-commerce, ouvrages entreprise, management, emploi, ressources humaines, s&eacute;lection de
livre, immobilier, urbanisme, sciences, developpement personnel, bricolage, droit, graphisme,
audiovisuel, Livre, Référence sur le net, S. Saporta, Les nouvelles méthodes gagnantes,
marketing web" />
  <meta name="description" content="Librairie Eyrolles, vente de livres spécialis&eacute;s en
informatique, livres entreprise, emploi, livres management, livres developpement personnel, livres
bricolage, livres droit, livres graphisme, audiovisuel s&eacute;lection de livre d immobilier et
urbanisme" />
  <style type="text/css" media="screen">
@import url("/layout/librairie/css/default.css");
@import url("/layout/librairie/css/acc/style.css");
@import url("/layout/librairie/css/livre.css");
</style>
  <script type="text/javascript" src="/layout/librairie/js/commun.js"></script>
  <link rel="shortcut icon" type="images/x-icon" href="/favicon.ico" />
</head>
```


Annexe 13 : capture d'écran d'une page produit Lavoisier.fr



Annexe 14 : code source HTML d'une page produit Lavoisier.fr

```
<html>
<head>
<title>LAVOISIER</title>
</head>
<frameset rows="109,*" border="0" framespacing="0" frameborder="NO">
  <frame src="/fr/livres/hautlivres.asp" name="hautlivres" scrolling="NO" frameborder="NO"
topmargin=0 leftmargin=0 marginwidth="0" marginheight="0">
  <frame
src="detail.asp?texte=r%E9f%E9rencement&select=motcle&exact=on&support=NULL&from="
name="baslivres" leftmargin=0 marginwidth="0" frameborder="NO">
</frameset>
<noframes><body bgcolor="#FFFFFF">
</body></noframes>
</html>
```

Annexe 15 : positionnement des sites étudiés sur les moteurs de recherche

	Alapage	Amazon	Chapitre	Fnac	Dunod	Eyrolles	Lavoisier
1	Yahoo France - Google France - AltaVista France - MSN France -	Yahoo France 1 Google France 1 AltaVista France 1 MSN France 1	Yahoo France - Google France - AltaVista France - MSN France -	Yahoo France 2 Google France 15 AltaVista France 2 MSN France -	Dunod ne vend pas cet ouvrage	Yahoo France - Google France 3 AltaVista France - MSN France 3	Yahoo France - Google France - AltaVista France - MSN France -
2	Yahoo France - Google France - AltaVista France - MSN France -	Yahoo France 1 Google France 1 AltaVista France 1 MSN France -	Yahoo France - Google France - AltaVista France - MSN France -	La Fnac ne vend pas cet ouvrage	Yahoo France 7 Google France 3 AltaVista France 7 MSN France 3	Yahoo France - Google France - AltaVista France - MSN France -	Yahoo France - Google France - AltaVista France - MSN France -
3	Yahoo France - Google France - AltaVista France - MSN France -	Yahoo France 1 Google France 1 AltaVista France 1 MSN France -	Yahoo France - Google France - AltaVista France - MSN France -	Yahoo France 4 Google France - AltaVista France - MSN France -	Dunod ne vend pas cet ouvrage	Yahoo France - Google France 4 AltaVista France - MSN France -	Yahoo France - Google France - AltaVista France - MSN France -

Source : Outiref.com .- Consulté le 14 septembre 2005 et vérification manuelle sur chacun des moteurs

Annexe 16 : texte visible des pages produits

Site	Analyse
Alapage	Le texte visible de votre page fait 2367 caractères, et 358 mots. (Poids total de la page : 51306 caractères)
Amazon	Le texte visible de votre page fait 4787 caractères, et 775 mots. (Poids total de la page : 50940 caractères)
Chapitre	Le texte visible de votre page fait 82 caractères, et 15 mots. (Poids total de la page : 2445 caractères)
Fnac	Le texte visible de votre page fait 2942 caractères, et 445 mots. (Poids total de la page : 43207 caractères)
Dunod	Le texte visible de votre page fait 1385 caractères, et 211 mots. (Poids total de la page : 18186 caractères)
Eyrolles	Le texte visible de votre page fait 5972 caractères, et 790 mots. (Poids total de la page : 33856 caractères)
Lavoisier	Le texte visible de votre page fait 1 caractères, et 0 mots. (Poids total de la page : 499 caractères)

* Livre de Sandrine Saporta, *Référencement sur le net, les techniques gagnantes*, à l'exception de Dunod : Hardy J.-M., *Check-list pour réussir son site web*.

Source : Outiref.com. – Consulté le 8 septembre 2005

Annexe 17 : densité des mots trouvés dans le texte visible des pages produits

Alapage	Mots clés uniques : 165 Expression de 2 mots-clés : 201 Expression de 3 mots-clés : 215	10 livres (4,22%) 4 référencement (1,69%) 3 livre (1,27%) 4 référencement net (1,69%) 2 référencement net référencement (0,85%) 2 net référencement net (0,85%)
Amazon	Mots clés uniques : 276 Expression de 2 mots-clés : 406 Expression de 3 mots-clés : 452	16 livres (3,29%) 10 référencement (2,05%) 6 saporta (1,23%) 5 livre (1,03%) 8 référencement net (1,65%) 6 Sandrine Saporta (1,23%)
Chapitre	Mots clés uniques : 10 Expression de 2 mots-clés : 9 Expression de 3 mots-clés : 8	
Fnac	Mots clés uniques : 238 Expression de 2 mots-clés : 309 Expression de 3 mots-clés : 318	9 référencement (2,77%) 4 saporta (1,23%) 4 livre (1,23%) 3 livres (0,92%) 2 net saporta (0,62%) 2 référencement net saporta (0,62%)
Dunod	Mots clés uniques : 112 Expression de 2 mots-clés : 145 Expression de 3 mots-clés : 153	8 web (4,79%) 7 site (4,19%) 5 livre (2,99%) 3 check-list (1,80%) 5 site web (3,01%) 4 réussir site (2,41%) 3 check-list réussir (1,81%) 4 réussir site web (2,42%) 3 check-list réussir site (1,82%) 2 livre check-list réussir (1,21%)
Eyrolles	Mots clés uniques : 349 Expression de 2 mots-clés : 511 Expression de 3 mots-clés : 553	11 référencement (1,82%) 9 livre (1,49%) 9 livres (1,49%) 4 techniques référencement (0,66%)
Lavoisier	Mots clés uniques : 1	lavoisier (100%)

Source : Outiref.com .- Consulté le 8 septembre 2005

Annexe 18 : liens sponsorisés sur certains mots-clés

MIVA	Google	Overture
10 premiers liens sponsorisés sur le mot « livre »		
2. alapage 0,12€ 3. chapitre 0,11€	1. chapitre 2. fnac	1. chapitre 0,36€ 2. alapage 0,31€
10 premiers liens sponsorisés sur le mot « livres »		
1. chapitre 0,13€ 2. alapage 0,11€	1. chapitre 2. fnac	1. chapitre 0,36€ 2. alapage 0,31€
10 premiers liens sponsorisés sur le mot « librairie en ligne »		
1. alapage 0,11€ 2. chapitre 0,10€	1. amazon	1. alapage 0,25€ 2. chapitre 0,24€
10 premiers liens sponsorisés sur le mot « cyberlibrairie »		
1. chapitre 0,10€		
10 premiers liens sponsorisés sur le mot « vente livres »		
1.alapage 0,12€ 2.chapitre 0,11€	1.amazon 2.chapitre	2.alapage 0,23€ 3.chapitre 0,20€
10 premiers liens sponsorisés sur le mot « achat livres »		
1.chapitre 0,13€ 2.alapage 0,11€	1.amazon 2.chapitre	1.alapage 0,26€ 3. chapitre 0,20€
10 premiers liens sponsorisés sur le mot « livre informatique »		
1.chapitre 0,11€	1.amazon 3.chapitre 10.fnac	1.chapitre 0,20€ 2.alapage 0,16€
10 premiers liens sponsorisés sur le mot « livre internet »		
1. chapitre 0,10€	1.fnac 2.amazon 3.chapitre 7.chapitre	1.chapitre 0,20€

Source : Moklic, Miva et Overture .- Consultés le 9 septembre 2005 et le 15 septembre 2005

INDEX ALPHABETIQUE

A

Abondance · 38
 accès · 59
 achat de mots-clés · 49, 89, 90, 93
 acquisition clients · 52
Active Server Pages · Voir ASP
 adresse IP · 56, 58
 AdWords · 37, 89
 affiliation · 60, 63, 90, 93
 affilié · 63, 90, 91
 affilié · 63, 91
 AFNOR · 20
 agent · Voir robot
 Alapage · 13, 68, 69, 71, 73, 76, 77, 78, 79, 82, 83, 84, 89, 90, 91, 93
 algorithme · 27, 28, 48, 50
 algorithme de classement · 30, 36, 88
 algorithme de pertinence · 25
 ALT · 25, 36, 82, 92, 93
 AltaVista · 29, 43, 72
 Amazon · 13, 63, 68, 69, 71, 72, 73, 74, 76, 77, 78, 79, 82, 83, 86, 87, 89, 90, 91, 92, 93
 annonceur · 42, 43, 44, 45, 52, 53
 annuaire · 13, 16, 18, 23, 32, 35, 40, 47, 48, 53
 Aol · 85
 Apache · 71
 applet java · 70
 araignée · Voir robot
 arborescence · 36, 40
 ASP · 70, 71
 aspirateur · Voir robot
 audience · 56, 57

B

backlink · 27
 balise <META> · 20, 22, 23, 36, 79
 balise <TITLE> · 22, 75, 79, 88
 balise meta Description · 23, 76, 77
 balise meta Keywords · 23, 76, 79
 bande passante · 59
 bandeau · 43, 64, 65, 91

bannière · 64, 90, 91
 base de données · 18, 28, 29, 34, 43, 45, 46, 70, 71, 83
 bibliothèque · 24, 56
 Brioude Internet · 85
 bruit · 32, 37

C

cadre · 93 Voir frame
 Cascading Style Sheet · 80
 CGI · 30, 39
 Chapitre · 69, 71, 73, 76, 77, 79, 81, 82, 84, 89, 90, 93
 Cibleclick · 90
 classement de pertinence · 87
 clic · 26, 41, 42, 43, 44, 45, 46, 89
cloaking · 48
clustering · 28
Common Gateway Interface · Voir CGI
 contrat de référencement · 50
 cookie · 58
 coût d'acquisition client · 43
 coût par clic · 43, 44, 45
 coût pour mille impressions · 44
 CPC · Voir coût par clic
 CPM · Voir coût pour mille impressions
crawler · Voir robot
 critère « in-page » · 22
 critère « off-page » · 26
 critère de classement · 28, 42
 critère de pertinence · 22, 25, 36, 75, 79, 92, 95
 cyberlibrairie · 13, 67, 68, 69, 89
 cybermarketing · 63
 cybermarketing direct · 64
 cybermarketing tribal · 64
 cybermarketing viral · 64

D

deep crawler · 19
 densité de mots-clés · 84
 descripteur · 37
 DirectHit · 28

Dmoz · Voir Open Directory Project
DNS · 39

Domain Name Server · Voir DNS

Dublin Core · 23, 24, 80

Dunod · 69, 71, 73, 74, 76, 77, 78, 79, 80, 81, 82, 83, 84, 86, 87, 90, 93

E

éditeur · 42, 43, 45, 61, 69, 70

enchères · 43, 44, 49, 53, 61, 89

ergonomie · 58

Exalead · 28, 31

Excite · 31, 34

Eyrolles · 70, 71, 73, 76, 77, 78, 79, 80, 81, 82, 83, 86, 90, 91, 93

F

facturation à la performance · 43

fichier inverse · 20

fil RSS · 64, 91

Flash · 29, 39, 81, 82, 92

flux XML · 45

Fnac · 13, 69, 71, 72, 73, 76, 77, 78, 79, 80, 83, 84, 86, 87, 89, 90, 93

formulaire · 19, 30, 40, 47, 72

forum · 65

frame · 29, 39, 81, 82, 83, 84

France Télécom · 68

fresh crawler · 19

full crawler · Voir deep crawler

full text · Voir texte intégral

G

Google · 17, 18, 19, 23, 27, 28, 29, 32, 42, 43, 44, 45, 46, 49, 72, 81, 85, 86, 87, 88, 89, 92, 93

Google AdSense · 44

Google Dance · 28

groupe de discussion · 65

guide · Voir annuaire

H

hit · 57, 58

HP-UX · 71

HTML · 26, 67, 75, 76, 92

HTTP · 23

Hypertext Preprocessor · Voir PHP

I

ID · 25 Voir indice de densité

inclusion dynamique · Voir flux XML

inclusion payante d'URL · 45

index · 20, 25, 27, 30, 35, 41, 45, 92

index inversé · 20

index principal · 20

indexation · 13, 18, 19, 20, 21, 35, 36, 38, 45, 46, 50, 92, 93

indexation automatique · 19, 31, 35, 61

indexation documentaire · 20, 21, 35

indexation en texte intégral · 21

indexation informatique · 20, 35

indice de clic · 26

indice de densité · 25, 26, 84, 92

indice de popularité · 26, 87, 91

interface de recherche · 18

interrogation simple · 31

IP · Voir indice de popularité

IPP · 26 Voir indice de popularité

J

Javascript · 29, 39, 70

L

langage documentaire · 21

Lavoisier · 70, 71, 73, 76, 77, 79, 81, 82, 84, 90, 93

lien commercial · Voir lien sponsorisé

lien contextuel · 14, 44

lien promotionnel · 43, 44, 50

lien publicitaire · 44, 50

lien sponsorisé · 14, 31, 41, 43, 44, 46, 48, 53, 61, 64, 89

Linux · 71

liste noire · 49
log · 56, 57
 Lycos France · 16

M

maison d'édition · 13, 67, 68, 69, 89, 93
 marketing · 13, 38, 41, 43, 50, 52, 54, 56, 59, 61, 63, 64, 79, 89, 95
 marketing adapté · 55
 marketing concentré · 54
 marketing différencié · 55
 marketing indifférencié · 54
 Mediamétrie · 17, 69, 91
 mesure côté client · 56
 mesure côté serveur · 56
 mesure d'audience · 56
 meta HTTP-EQUIV · 23, 80
 meta name · 23, 79
meta tag · Voir balise <META>
 métadonnée · 18, 22, 23, 24, 67, 76, 79, 80, 88, 93
 méta-moteur · 17
 Miva · 42, 89
 Moklic · 89
 mot primaire · 37
 mot secondaire · 37
 mot-clé · 18, 20, 21, 22, 23, 25, 26, 30, 34, 36, 37, 38, 41, 42, 43, 44, 46, 48, 49, 50, 52, 53, 54, 57, 58, 59, 60, 61, 62, 76, 79, 81, 84, 88, 89, 92, 93
 moteur d'indexation · 19
 moteur d'interrogation · 30
 moteur de recherche · 13, 14, 16, 17, 18, 20, 23, 24, 25, 29, 30, 32, 34, 35, 36, 41, 42, 45, 48, 53, 61, 72, 73, 75, 76, 80, 81, 82, 83, 85, 88, 90, 91, 92, 95
 moteur de recherche interne · 59
 MSN · 17, 43, 72, 85
 MySQL · 71

N

navigateur · 16, 23, 30, 39, 58, 60, 67, 70, 80, 134
netlinking · 65

Netscape · 71
 netsurfeur · 16, 40
 nom de domaine · 16, 26
 Nomade · 16, 44
 norme PICS-LABEL · 80
 Northernlight · 28

O

Open Directory Project · 16, 18
 opérateur booléen · 30, 31
 optimisation · 14, 37, 46, 61, 67, 72, 74, 92, 93, 95
 Oracle · 71
 outil d'analyse · 55, 56, 58
 outil de recherche · 2, 13, 16, 17, 34, 35, 36, 37, 40, 41, 42, 47, 48, 50, 53, 57, 58, 59, 60, 61, 62, 73, 85
 Outiref · 73, 74
 Outrider · 32
 Overture · 42, 49, 89

P

page alias · 48
 page de résultats · 31
 page dynamique · 46, 83
 page interstitielle · 64
 page orpheline · 29
 page satellite · 46, 48
 page statique · 46
 page vue · 56, 57, 58, 59
 PageRank© · 27, 28, 86, 87
paid inclusion · 46, 53 Voir flux XML
 parrainage · 64
 partenariat · 60, 63, 90
pay inclusion · 45
 Pay Per Clic · Voir coût par clic
 Perl-Pro · 85
 pertinence · 18, 22, 24, 49
 PFI · Voir Pay Inclusion
 PHP · 70, 71
 Pinault Printemps Redoute · 69
Platform for Internet Content Selection · Voir norme PICS-LABEL
 popularité · 27, 48, 59, 87, 88, 91, 93

pop-up · 64
 positionnement · 13, 14, 22, 27, 28, 34, 36,
 37, 41, 42, 44, 45, 46, 48, 49, 50, 52, 53,
 57, 58, 61, 62, 67, 72, 73, 74, 85, 91, 92,
 95
 positionnement payant · 42
 PostgreSQL · 71
 PPC · Voir Pay Per Clic
 PR · 27, 86, 87, 88 Voir PageRank
 promotion · 59, 65, 89, 93
 promotion indirecte · 65
 publicité · 42, 44, 60, 91
 publicité en ligne · 64

R

ranking · 22, 27
 rapport de positionnement · 62
Really Simple Syndication · Voir fil RSS
 recherche avancée · 30
 recherche ciblée · 30
 recherche simple · 30, 72
 réécriture d'URL · 46, 71
 référencement · 13, 14, 32, 34, 35, 36, 37,
 38, 39, 40, 41, 45, 46, 47, 48, 50, 52, 54,
 55, 56, 57, 59, 60, 61, 62, 63, 67, 72, 76,
 81, 82, 92
 référencement automatique · 47
 référencement manuel · 47
 référencement naturel · 41, 50
 référencement par indexation · 41
 référencement par positionnement · 41, 42
 référencement payant · 42
 régie · 42, 89
 réindexation · 61
 rentabilité · 56
 répertoire · 16 Voir annuaire
 requête · 16, 27, 29, 30, 31, 32, 37, 41, 43,
 44, 52, 56, 72, 73, 74, 86, 95
 retour sur investissement · 43, 46, 60, 62
Rich Site Summary · Voir fil RSS
 robot · 16, 18, 19, 20, 25, 31, 35, 36, 39,
 45, 46, 48, 56, 79

S

search engine marketing · 50
search marketing · 95
 segmentation du public · 54
 serveur web · 46, 56, 57, 59
 SGBD · 71
 site dynamique · 45, 70
site-centric · 56 Voir mesure côté serveur
sniffer · 57
 société de référencement · 42
 soumission · 40, 46, 48
 soumission gratuite · 40, 48
 soumission payante · 40
spamdexing · 26, 28, 39, 46, 48
spider · Voir robot
sponsoring · Voir parrainage
 statistique · 56, 59
 stratégie de recherche · 31, 53
 stratégie marketing · 13, 14, 93
 Sybase · 71
 syndication · 63, 64, 91
 système d'exploitation · 58
 système de gestion de bases de données ·
 Voir SGBD

T

taux de transformation · 44
 texte intégral · 19, 20, 21, 30, 32
 trafic · 13, 19, 27, 34, 35, 37, 43, 44, 46,
 56, 57, 58, 61, 62, 64, 85, 86, 87, 89
 trafic qualifié · 44, 52
 tri des résultats · 22
 tri par catégories · 28
 tri par mesure d'audience · 28
 tri par pertinence · 27
 tri par popularité · 27
trust feed · Voir flux XML

U

uniterme · 31
 Unix · 71
 Unix dbm · 71
 URL · 25, 39, 40, 46, 60, 70, 71, 83, 88, 93

URL rewriting · Voir réécriture d'URL
user-centric · Voir mesure côté client

V

vedette-matière · 37
veille · 50, 61
visibilité · 13, 14, 32, 34, 40, 41, 43, 44,
45, 50, 52, 61, 63, 67, 85, 86, 88, 89, 92,
93, 95
Visiref · 85
visite · 34, 57, 58, 59, 60
Voila · 16, 50, 85

W

W3C · 24

Wanadoo · 68
wanderer · Voir robot
web marketing · 43, 50
webmaster · 46 Voir webmestre
webmestre · 28, 79
Webring · 64

X

Xiti · 17
XML · 45, 64
XML PI · Voir flux XML

Y

Yahoo · 16, 17, 18, 29, 32, 40, 43, 44, 72,
85, 87, 92