



HAL
open science

Analyse des techniques d'indexation des pages Web et de référencement des sites sur Internet et gestion du site

Lille 3 Jeunesse

Jonathan Marie

► To cite this version:

Jonathan Marie. Analyse des techniques d'indexation des pages Web et de référencement des sites sur Internet et gestion du site Lille 3 Jeunesse. Sciences de l'information et de la communication. 2004. dumas-01552403

HAL Id: dumas-01552403

<https://dumas.ccsd.cnrs.fr/dumas-01552403>

Submitted on 12 Dec 2017

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.

04 S EDU

MAR

Jonathan MARIE

Master SID I (Sciences de l'Information et du Document)

Mémoire de stage et de recherche

Mission effectuée du 1^{er} novembre 2003 au 30 juin 2004
à la Bibliothèque des Bibliothèques de l'Université Charles-De-Gaulle Lille 3

**Analyse des techniques d'indexation des pages Web et de
référencement des sites sur Internet et gestion du site
*Lille3 Jeunesse***

Sous la direction de :

M. Christian LOOCK & M. Ismaïl TIMIMI (responsables universitaires)

Mme Annette BEGUIN (tuteur professionnel)

Soutenu le 22 septembre 2004 à l'UFR IDIST
Université Charles-de-Gaulle, Lille 3
BP 49, 59650 Villeneuve d'Ascq

Année universitaire 2003/2004



2005* 312
04 S EDU MAR

REMERCIEMENTS.....	- 1 -
INTRODUCTION.....	- 2 -
I. ANALYSE DES TECHNIQUES D'INDEXATION ET DE RÉFÉRENCIEMENT.....	- 4 -
A. L'INDEXATION DES PAGES WEB.....	- 4 -
1. <i>État de l'information sur l'Internet</i>	- 4 -
a) L'explosion de l'information.....	- 4 -
b) Quel traitement appliquer à l'information en ligne ?.....	- 5 -
2. <i>Tentative d'application d'une norme aux métadonnées</i>	- 6 -
a) Les métadonnées.....	- 6 -
b) L'Initiative de Métadonnées du Dublin Core.....	- 7 -
c) Le Web sémantique.....	- 9 -
d) Les limites à la normalisation.....	- 9 -
3. <i>Le choix des mots-clés et les techniques d'indexation</i>	- 10 -
a) Préparer un site au référencement.....	- 11 -
b) Méthodes pour le choix et l'insertion des mots-clés.....	- 12 -
(1) Le choix des mots-clés.....	- 12 -
(2) L'insertion des mots-clés.....	- 13 -
(3) Les robots.....	- 15 -
(4) Les techniques de fraude.....	- 15 -
B. LE RÉFÉRENCIEMENT.....	- 16 -
1. <i>Intérêt du référencement</i>	- 16 -
2. <i>Aperçu des méthodes de recherche d'information sur le Web</i>	- 17 -
3. <i>Le Web invisible</i>	- 18 -
4. <i>Méthodes pour le référencement d'un site Internet</i>	- 19 -
5. <i>Les obstacles au référencement</i>	- 21 -
6. <i>D'autres techniques de promotion d'un site Internet</i>	- 21 -
7. <i>Comment maintenir le niveau de popularité</i>	- 23 -
II. LE SITE LILLE3 JEUNESSE.....	- 24 -
A. PRÉSENTATION DU CADRE DU STAGE.....	- 25 -
1. <i>La littérature jeunesse à l'université</i>	- 25 -
2. <i>Le site Lille3 Jeunesse</i>	- 25 -
a) L'historique.....	- 25 -
b) Le contenu.....	- 26 -
B. TÂCHES DE NATURE JOURNALISTIQUE.....	- 27 -
1. <i>La 8^{ème} rencontre professionnelle à l'UFM d'Arras</i>	- 27 -
2. <i>Les Parcours Professionnels Pour la Lecture Jeunesse</i>	- 27 -
C. « NOUS VOULONS LIRE » : LE JEU D'IMAGES.....	- 28 -
D. TÂCHES DE NATURE DOCUMENTAIRE.....	- 30 -
E. L'ENTRETIEN DU SITE.....	- 30 -
1. <i>Mes tâches</i>	- 30 -
a) La bibliographie thématique.....	- 31 -
b) Le Mag Garçons / Filles.....	- 31 -
c) La Vie des Livres.....	- 31 -
d) La navigabilité.....	- 33 -
2. <i>Tâches partagées avec Julien Siméoni</i>	- 33 -
F. APPLICATION DE LA PARTIE THÉORIQUE AU SITE LILLE3 JEUNESSE.....	- 34 -
1. <i>Le Dublin Core</i>	- 34 -
a) L'Éducation Nationale et le Dublin Core.....	- 35 -
b) La syntaxe des métadonnées.....	- 35 -
2. <i>Les mots-clés</i>	- 36 -
a) Le contexte.....	- 36 -
b) Composer la liste.....	- 36 -
c) Classer les mots-clés.....	- 37 -
d) L'insertion des mots-clés dans la page.....	- 37 -
e) Piloter les robots.....	- 37 -
3. <i>Le référencement du site</i>	- 37 -
a) Réviser le code HTML des frames.....	- 37 -
b) Le choix des outils de recherche adéquats.....	- 38 -
c) La phase de référencement.....	- 38 -
4. <i>Les autres techniques de promotion</i>	- 38 -
CONCLUSION.....	- 40 -
BIBLIOGRAPHIE.....	- 41 -
GLOSSAIRE.....	- 45 -
ANNEXES.....	- 49 -

Remerciements

A M. Christian Loock qui a su nous transmettre sa passion pour la littérature jeunesse et qui n'est jamais à court d'idées pour ce qui est de la faire connaître,

A Julien, Élodie, Coralie, Charlotte, Juliette, Nicolas qui ont osé se lancer dans cette « aventure » et qui ont formé un groupe des plus sympathiques,

A Marine Dormion qui n'a pas eu froid aux yeux de s'occuper seule d'un projet d'une telle envergure,

A M. Ismaïl Timimi qui m'a laissé libre dans mes réflexions,

A mes parents, Anne & Jean-Luc, pour leur extrême patience,

Et à Adeline pour son immense soutien et sans qui ce mémoire ne serait pas,

Merci...

Introduction

Travailler sur un site Internet pose d'emblée la question du référencement de ce site. En effet, le webmaster se trouve souvent tiraillé entre deux logiques contradictoires. Une première logique, plus commerciale, vise à « populariser » et faire obtenir son site le plus fréquemment possible comme réponse à la requête d'un usager en utilisant des mots-clés qui sont susceptibles d'être employés lors d'une recherche effectuée par un utilisateur. Ce faisant, le webmaster en oublie souvent la deuxième logique à visée plus documentaire, qu'il convient de mettre en place. La précision documentaire doit présider au choix des termes employés pour indexer les pages du site.

D'aucun peut facilement remarquer qu'un des travers les plus courants des webmasters est qu'ils se soucient avant tout de l'esthétique du site qu'ils gèrent, parfois au détriment de sa précision documentaire. Autrement dit, ils pensent d'abord en termes de « look ». Cependant, cette façon de faire répond à une attente très nette des usagers qui se montrent vite agacés lorsqu'ils se trouvent face à des pavés de texte d'une longueur décourageante ou lorsqu'ils doivent patienter plus que de raison pour pouvoir charger une page. Il est vrai que les universitaires constituent une exception à cette norme, ils s'avèrent souvent plus enclins à lire des pages assez conséquentes, mais ils ne sauraient être pris comme une généralité. Des désirs des usagers, il ressort que tout doit être rapide : un contenu doit être rapidement chargé et rapidement lu. Or, de nos jours, la création d'un site, si elle n'est pas encore sans difficulté, ne requiert plus des compétences pointues. De ce fait, le nombre de pages existantes sur le Web a explosé¹. L'information se retrouve alors noyée dans la masse, il est plus long et fastidieux d'obtenir une information pertinente.

Mais on a vu que le surfer aime la rapidité, il devient donc indispensable de rendre l'accès à un site aisé.

Le site Web Lille3 Jeunesse ne déroge pas à la règle. C'est pourquoi je me suis attaché cette année, en accord avec M. Loock, à essayer de trouver des solutions pour rendre le site plus accessible aux personnes dont la recherche présente une adéquation certaine avec ce que propose le contenu du site. Concrètement, l'objectif principal est d'être référencé en bonne place et de manière pertinente dans les moteurs de recherches et annuaires les plus connus (et donc les plus susceptibles d'être utilisés). Il s'agit de procéder à un référencement raisonné qui demande l'emploi judicieux de mots-clés choisis avec discernement.

Néanmoins, une petite difficulté subsiste avec le site Lille3 Jeunesse puisqu'il vaut mieux amorcer cette phase de référencement en amont de la création des pages, pendant la conception du site, ou au début de l'existence du site sur l'Internet². Ce n'était pas le cas avec le site dont je me suis occupé. En effet, la popularité n'était pas un objectif premier lors de la création. Aucune politique de communication n'avait été mise en place puisque le public intéressé était majoritairement universitaire. La prise d'ampleur rapide du site et l'évolution du public ont accentué la nécessité de le faire connaître des personnes intéressées.

La tâche qui m'a été confiée s'inscrit dans la politique d'optimisation pour les moteurs de recherche (l'OMR), qui elle-même constitue une partie de l'optimisation des sites Web (l'OSW). Cette politique demande l'utilisation pertinente de mots-clés disposés de façon réfléchie dans la page Web. Elle réclame aussi d'agir en complémentarité avec l'ergonomie du site : il ne sert à rien de créer un beau site si celui-ci ne peut être rapidement trouvé. De

¹ Selon B. FOENIX-RIOU, *Recherche et veille sur le web visible et invisible*, Paris : Bases Publications, 2001, p. 2-3.

² Selon SYS-COM, *Stratégie de test e-business*, Paris : Hermès Science, 2001, p. 65.

même, la rapidité doit être vue en parallèle avec la facilité pour l'utilisateur de retrouver le site, car « un site Web rapide comme l'éclair reste peu de choses si personne ne peut le découvrir »³.

Une dernière composante entre en ligne de compte pour l'indexation des pages de Lille3 Jeunesse, il s'agit d'un site universitaire duquel il émane une certaine image de sérieux et de véracité de l'information. Il faut maintenir cette autorité intellectuelle jusque dans la conception du site. Dans ce cas, il est légitime de s'interroger sur l'utilité éventuelle de l'emploi d'une des normes actuellement développées pour être appliquées aux sites Web.

A. B. KING, *Optimisation de sites Web*, Paris : CampusPress, 2003, p. 303

I. Analyse des techniques d'indexation et de référencement

L'inflation informationnelle actuelle qui existe sur l'Internet demande d'appliquer des techniques de plus en plus perfectionnées pour permettre à un site d'être repérable. Dans le contexte où des milliers de sites voient le jour quotidiennement, il convient d'être rigoureux dans les opérations d'indexation et de référencement pour acquérir une place honorable dans le classement des outils de recherche. Précisons dès à présent quelles sont les différences entre l'indexation et le référencement.

L'indexation est l'insertion dans le code d'une page Web de mots-clés adéquats destinés à décrire le contenu de cette page. Une bonne indexation ne se fait pas au hasard, elle reste un moyen primordial pour optimiser le score attribué par les robots des moteurs de recherche. Il existe quelques préceptes à suivre et des limites qu'il vaut mieux respecter. Le référencement vient en complément de l'indexation. Il consiste à proposer aux outils de recherche certaines pages d'un site Web. Cette opération est différenciée de celle de l'indexation par une volonté initiale du gestionnaire du site d'amener un moteur ou un annuaire sur son site. L'indexation par les moteurs de recherche étant faite de manière automatique, elle est sujette à des délais de plus en plus longs, à des aléas qui font que même bien réalisée par le webmaster elle peut ne pas être efficace. Le référencement est un moyen de pallier ces incertitudes en inscrivant l'indexation des pages dans une politique de promotion du site.

Une logique de performance s'étant installée très tôt sur Internet, il est facile de perdre de vue que l'indexation a une visée documentaire avant d'être un outil commercial. A ce titre, il peut sembler souhaitable de la soumettre à une certaine normalisation et dans la syntaxe du code de la page et dans le langage utilisé pour indexer. Actuellement, on n'en est qu'aux balbutiements d'une telle réglementation qui ne sera peut-être jamais possible. Des groupes de travaux s'organisent pour tenter d'instaurer un cadre normatif assez souple pour convenir à la diversité des internautes.

A. L'indexation des pages Web

1. État de l'information sur l'Internet

a) L'explosion de l'information

La situation actuelle sur Internet nécessiterait peut-être l'emploi d'une norme d'indexation structurée. En effet, la multitude des éditeurs de l'information et les difficultés pour les identifier de manière sûre (universités, centres de recherche, entreprises, mais également sites commerciaux, centres de documentation, bases de données payantes, agences de presse, et encore sites gouvernementaux, parapublics, institutionnels, sites portails,

associatifs, organisations privées diverses, forums de discussion, et pages privées). Cela engendre un manque de fiabilité de l'information a priori. Un filtrage mais surtout une qualification des sources sont fondamentaux. La validité de l'information peut être légitimement questionnée puisque très peu de ce qui est mis en ligne est contrôlé par une instance quelconque. Ce n'est pas le cas pour l'information papier qui subit un contrôle minimum, par l'éditeur le plus souvent⁴.

La quantité exceptionnelle de sujets traités s'avère être un autre problème de taille. Elle a pour conséquence l'augmentation du bruit liée à la polysémie et à l'homographie lors des recherches, augmentation renforcée par des présentations différentes dues aux différents points de vue qui peuvent coexister sur un même sujet. Cela est voué à perdurer car de plus en plus de domaines sont traités de manière de plus en plus conséquente. La masse informationnelle va continuer à augmenter. Les volumes d'information engendrés vont rapidement demander une évolution du traitement des données et des outils qui les traitent.

b) Quel traitement appliquer à l'information en ligne ?

Il semble évident que le traitement humain de l'information est préférable. Mais la croissance exponentielle de la masse informationnelle rend impossible ce seul traitement. La quantité à gérer est trop importante et elle ne cesse de croître, en outre la vitesse d'exécution est bien sûr en deçà de celle d'une indexation automatisée. Cependant, le traitement robotisé présente un inconvénient assez important, il n'est pour l'instant pas en mesure de dissocier les différents types d'information, à savoir les sources d'information primaire et les sources d'information secondaire qui se côtoient très fréquemment dans un même site, à l'instar des portails, des sites spécialisés qui proposent des résumés, des études comparatives, des liens vers d'autres sites, des références bibliographiques, des documents primaires, ... Seul l'humain peut traiter et analyser ces différents types d'information, pour enfin les gérer.

Il est néanmoins indispensable de procéder à une indexation « intelligente » des pages du World Wide Web par des robots de moteurs de recherche. Toutefois, cette indexation dite intelligente se fait sur un nivellement de l'information : c'est le principe de neutralité, les moteurs partent du principe que toutes les informations se valent et qu'elle méritent toutes de circuler au même plan. Cette méthode a l'avantage de structurer un minimum le flot d'information, mais il ne fait pas de réels choix cognitifs, pas plus qu'il ne hiérarchise la qualité de l'information ou ne propose plusieurs « chemins » selon des critères de validité solidement éprouvés.

En effet, les robots se basent essentiellement sur les contenus des balises métatags (les mots-clés) dont le choix est laissé à la discrétion des webmasters, et qui répondent donc généralement plutôt à une logique commerciale ou qui sont tout au moins sans logique documentaire. Un problème de clarté, de pertinence et d'honnêteté se pose donc⁵.

Internet est arrivé rapidement et s'est très vite développé obligeant de ce fait les informaticiens à créer dans l'urgence des techniques applicables au Web mais qui n'étaient pas du niveau de ce qui se faisait (et se fait encore) dans les autres domaines de la documentation (en particulier la recherche d'informations). Or, aujourd'hui, se positionner de manière visible dans l'Internet oblige le recours à des techniques multimédia mais aussi à des techniques d'ingénierie linguistique. Il semble difficile d'aller de l'avant sans allier les

⁴ Voir Annexe 2 : Contexte de la recherche d'informations dans un système documentaire classique et sur l'Internet.

⁵ Selon P. MATHIAS, Anatomie d'une abondance, in *Comprendre les usages de l'Internet*, Paris : Éditions Rue d'Ulm, 2001, p. 131

techniques actuelles de recherche sur le Web avec les techniques de documentation en général. Il faut que les travaux dans ces deux domaines respectifs soient menés conjointement pour espérer une progression.

2. Tentative d'application d'une norme aux métadonnées

a) Les métadonnées

La situation du Web en général amène à penser qu'il faudrait peut-être aller vers une structure plus marquée au moyen des métadonnées. En effet, les informations circulant sont lisibles par les robots des moteurs de recherche mais ils ne les comprennent pas plus qu'ils ne peuvent les interpréter. Les métadonnées permettent de donner un sens structuré et cognitif à l'information. Ce sont elles notamment qui vont permettre aux moteurs de recherche d'indexer les pages Web qui les contiennent.

Mais qu'est-ce qu'une métadonnée ? Littéralement, il s'agit d'une donnée sur une donnée. *Meta* vient du grec et signifie « avec, en parallèle, après ». On peut rapporter une acception plus récente, *meta* sert alors à exprimer quelque chose hors de nature. Les métadonnées forment un ensemble structuré d'informations servant à décrire une ressource quelconque. Cette dernière ne se trouve pas forcément sous une forme numérique, il peut s'agir par exemple d'un catalogue de bibliothèque sous une forme papier. Les métadonnées se constituent de mots-clés ou de textes libres et elles remplissent différents usages tels que la description, la gestion et la préservation du document auquel elles sont appliquées.

Les liens qui unissent une métadonnée à la ressource qu'elle décrit peuvent être de différentes natures : elle peut être séparée du contenu décrit, comme dans un catalogue de bibliothèque où la fiche descriptive d'un ouvrage est physiquement séparée de celui-ci par exemple, ou être incluse directement dans la ressource, comme c'est le cas pour une donnée numérique où la métadonnée est inscrite directement dans le code.

Cependant, on a assisté au début des années quatre-vingt-dix à une explosion de l'utilisation pas toujours rigoureuse des métadonnées sur Internet. S'ensuivit une surcharge d'information qui a nécessité une tentative pour établir des normes descriptives visant à améliorer la recherche de ressources pertinentes. Ces normes doivent faciliter la recherche d'information par une meilleure description, un meilleur classement des pages par les moteurs de recherche. Elles doivent en outre favoriser l'interopérabilité (c'est-à-dire le partage et l'échange d'informations), faciliter la gestion et l'archivage ainsi que la gestion et la protection des droits d'auteurs et enfin l'authentification des documents.

Les métadonnées sont contenues dans le code par les métabalises. Ces deux termes se confondent souvent dans la littérature, métadonnée étant le plus usité et désignant parfois la métabalise par métonymie. Les métabalises sont incluses dans le code d'un document numérique dans la section *En-Tête* (ou *Head*). Elles sont utilisées pour décrire le contenu de la page. Comme il vient d'être dit, les métabalises contiennent les métadonnées et n'ont pas pour vocation d'être visibles par les usagers des sites. Leur fonction essentielle est de permettre à un site d'être retrouvé sur le Web et d'amplifier la pertinence d'une page Web au moyen de descripteurs appropriés. Cependant, seules 21% des pages Web ont des métabalises renseignées de manière correcte.

Un archivage efficace passe par l'utilisation de métadonnées (dont l'ensemble peut être comparé à une fiche de bibliothèque électronique). Si l'on poursuit la comparaison, on s'aperçoit que rédiger une fiche bibliographique réclame l'emploi de formats standards. Il faut

donc appliquer cette règle aux métadonnées et essayer de leurs trouver des standards qui leurs sont propres. Cette standardisation est rendue d'autant plus nécessaire que l'indexation sur le Web n'est pas du seul fait de bibliothécaires ou de professionnels mais peut être réalisée par n'importe quel producteur de ressources (que ce soit un concepteur de sites chevronné ou un simple utilisateur de l'Internet qui aurait créé sa page personnelle). Il est évident que tous les créateurs de site n'évoluent pas dans le milieu de la documentation ou ne sont simplement pas familiarisés avec les préceptes qu'il dispense ou encore ne se sentent pas concernés par des enjeux documentaires. C'est pourquoi il paraît clair que la standardisation doit être mise au niveau de l'ensemble des internautes et ne doit pas être compréhensible des seuls professionnels de l'information. Cela peut sembler une tâche ardue à accomplir dans de telles conditions mais la standardisation s'avère nécessaire. Effectivement, si l'on poursuit la comparaison avec les bibliothèques entamée auparavant, on sait qu'un ouvrage mal catalogué est perdu pour l'utilisateur qui ne sera pas en mesure de le retrouver lors d'une recherche même si sa requête est formulée correctement. Il va sans dire qu'un service de prêt entre bibliothèques est hors de propos sans le respect de normes qui permettent à chaque établissement d'avoir les mêmes bases pour un catalogue collectif et raisonné. C'est d'autant plus vrai avec l'Internet où des métadonnées non renseignées ou attribuées sauvagement rendront l'interopérabilité impossible entre différentes collectivités qui ne pourraient s'adapter à toutes les façons d'indexer les documents.

Toutefois, l'élaboration de normes est actuellement confrontée à une prolifération de standards. Plusieurs standards pour les métadonnées voient simultanément le jour mais ont une orientation différente selon le « métier » (c'est-à-dire que chaque domaine a ses propres normes). La difficulté majeure n'est pas tant la prolifération des standards que la non interopérabilité de ceux-ci. Chaque standard a ses propres normes et ne peut être utilisé en dehors du « métier » pour lequel il a été conçu. Il faut donc plutôt chercher à créer une structure et une nomenclature minimales qui seraient applicables à n'importe quel domaine.

Beaucoup d'acteurs (pour la plupart issus des domaines de la documentation et du e-learning) se sont penchés sur ce problème : l'Aviation Industry Computer-based training Committee (AICC), le Dublin Core, l'Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE), l'Instructional Management Systems (IMS), le Sharable Content Object Reference Model (SCORM), l'Institute of Electrical and Electronics Engineers (IEEE), ...

b) L'Initiative de Métadonnées du Dublin Core

Nous allons étudier plus précisément l'Initiative de Métadonnées du Dublin Core (IMDC). Cette initiative s'est faite sous l'impulsion d'un certain nombre d'équipes impliquées dans la sémantique du Web et qui ont organisé un atelier au siège de l'OCLC (Online Computer Library Centre) à Dublin, dans l'Ohio, États-unis, en 1995. Leurs membres appartiennent surtout au NCSA (National Centre for Supercomputing Applications) et à l'OCLC. Une liste précise de 15 métadonnées communes à diverses communautés fut établie⁶. Elle a trait au contenu (*Couverture, Description, Sujet, Source, Titre, Type, Relation*), à la propriété intellectuelle (*Créateur, Collaborateur, Éditeur, Droits*) et à l'instanciation (*Date, Format, Identifiant, Langue*); en Dublin Core, l'instanciation représente « une occurrence spécifique d'une source d'information »⁷. Par exemple, un livre traduit est

⁶ Voir Annexe 4 : Champs de métadonnées du Dublin Core.

⁷ TEASDALE Guy, *Guide d'utilisation du Dublin Core*, <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm> [en ligne].

l'instanciation en français d'un livre original anglais. Il pourrait également y avoir une instanciation différente si le format change, ce qui modifierait d'autant la notice Dublin Core. Le Dublin Core ne dit pas comment représenter ces métadonnées dans la pratique, ce qui conduit à l'utilisation de plusieurs représentations. Cependant, deux conventions ont établi la syntaxe du Dublin Core : d'un côté, la syntaxe d'étiquettes HTML Méta (simplement placer les balises dans la zone *Head* du code) et de l'autre, le Resource Description Framework (RDF) (en français « le cadre de définition des ressources »). C'est le HTML Méta qui est le plus utilisé jusqu'à présent mais le RDF présente l'avantage d'être une expression du langage XML, plus normatif dans la façon d'écrire les données. La lisibilité des métadonnées par la machine est plus aisée avec le RDF.

Les métadonnées dans le Dublin Core sont appelées « éléments » dont la signification pour certains d'entre eux peut être précisés grâce à des « raffinements ». Ceux-ci sont facultatifs et restreignent la signification des éléments sans pour autant la changer fondamentalement. Les métadonnées s'expriment dans un format libre ou se conforment à un format bien défini⁸. Le Dublin Core s'appuie d'ailleurs sur certains formats qui font autorité (pour la date, le format est défini par la norme ISO 8601, pour la langue, le format est défini par la norme ISO RCF1766, pour l'identifiant on peut utiliser l'ISBN par exemple, etc.).

Si l'élaboration du Dublin Core a été effectuée par des professionnels de l'information, ses éléments ont été conçus de la manière la plus restreinte et simple possible afin de laisser aux non spécialistes la possibilité de s'en servir malgré tout. D'autre part, un souci d'objectivité a présidé à sa conception. Des facilités appréciables caractérisent le Dublin Core : sa gestion est assez simple, elle utilise une sémantique communément comprise, il a une envergure internationale, il est extensible et applicable à presque tous les formats de fichiers à condition que ceux-ci possèdent des métadonnées interprétables à la fois par les moteurs de recherche et par les humains⁹. Le Dublin Core favorise en outre l'interopérabilité sémantique et le multilinguisme. La participation de représentants de presque tous les continents au moment de sa création et l'appui du World Wide Web Consortium (W3C) place le Dublin Core en bonne position dans les standards les plus utilisés, il reçoit même les recommandations du groupe IETF, organisme qui fédère les groupes de recherche travaillant sur les technologies et les protocoles de l'Internet. De plus, il est en cours de normalisation par NISO (National Information Standards Organisation) aux États-unis et le Centre for European Standardisation en Europe. Cette technique est utilisée entre autres au sein des normes connexes par Open E-Book Publication Structure 1.0 (application à la publication de livres électroniques).

Le Dublin Core présente quand même des limites. Il ne prétend aucunement répondre aux besoins et à la complexité de tous les « métiers » et de tous les formats et doit d'ailleurs être généralement complété par d'autres schémas de métadonnées. De même, l'utilisation non réglementée du Dublin Core a compromis l'interopérabilité sémantique puisque l'interprétation de la définition des éléments constitutifs de cette normalisation est laissée à l'appréciation des utilisateurs de la norme.

Le Dublin Core constitue une première étape vers le Web sémantique qui est ce vers quoi les professionnels de l'information aimeraient voir le Web actuel évoluer.

⁸ Voir Annexe 5 : Formats de données conseillés par le Dublin Core.

⁹ Voir Annexe 3 : Utilisation de la syntaxe HTML avec le Dublin Core.

c) Le Web sémantique

Aujourd'hui le Web est composé de liens simples et universels mais qui ne sont sémantiquement pas ou peu structurants et structurés. On constate actuellement que les métadonnées sont peu ou mal utilisées. Certes, les moteurs de recherche sont de plus en plus sophistiqués mais ils demeurent imparfaits, d'où la nécessité d'instaurer des liens riches, d'utiliser des métadonnées structurées qui seraient sûres et signifiantes. Une première réponse a été apportée avec la conception d'un Web sémantique. Il s'agit d'une vision d'un Web suffisamment structuré pour permettre d'automatiser le traitement des données et de les intégrer et les réutiliser au travers de nombreuses applications¹⁰. Le Web sémantique permettrait d'intervenir à trois niveaux différents : au niveau des ressources : on veut pouvoir les nommer à l'aide de descripteurs universels – on parle des URI (Uniform Resource Identifier, identifiant uniforme de ressource) - et pouvoir les structurer à l'aide de schémas comme le DTD (Document Type Definition) ou encore les Schemas XML ; au niveau des métadonnées : on veut leur fournir un cadre comme le RDF et créer des ontologies destinées à définir les concepts liés à un terme. Les ontologies sont la principale lacune pour rendre le Web sémantique réel et elles se caractérisent par la difficulté de leur conception. Le niveau de la recherche est le dernier niveau, il faut qu'elle soit assurée par des agents intelligents capables de se servir des ontologies. Il serait en outre appréciable de parvenir à instaurer définitivement un système d'authentification des documents par un certificat, une signature, etc. Pour faire face à toutes ces exigences pour accéder au Web sémantique, une série de standards sont actuellement en cours d'élaboration mais les ontologies constituent un obstacle que les chercheurs n'arrivent aujourd'hui pas à surmonter. Il faut bien sûr faire concorder les technologies utilisées pour la recherche avec celles de l'indexation et favoriser l'emploi du RDF (basé sur les métadonnées) ou le Topic Maps (basé sur les réseaux sémantiques).

Cependant, si l'on fait un état des lieux de la façon d'indexer actuelle, force est de constater que la volonté de créer un Web sémantique restera une pure utopie. En effet, les métadonnées qui doivent être renseignées pour permettre l'existence de ce type de Web ne le sont pas ou sont remplies de manière peu fiable. Cela s'explique par le fait que les concepteurs de pages sont en général peu rigoureux, subjectifs même sans le vouloir dans l'attribution des descripteurs ou sont motivés par des objectifs purement commerciaux et vont donc attribuer des métadonnées dans l'unique but de faire de l'audience. Il reste malaisé de décrire objectivement son site, les schémas destinés à normaliser ne sont eux-mêmes jamais neutres, il n'y a pas de consensus général sur les normes à mettre en vigueur universellement et enfin il serait fallacieux de dire que l'on peut tout décrire d'une seule manière.

d) Les limites à la normalisation

Même si l'on considère que les métadonnées sont peu utilisées, leur nombre considérable (quand elles le sont) font que les moteurs de recherche les considèrent avant tout comme des générateurs d'abus (à l'exemple du *spamdexing* et autres), ce qui tient également à la manière dont les métadonnées sont renseignées. Cela conduit malheureusement à une prise en compte de plus en plus minime des métadonnées lors des référencements. Il est donc légitime de s'interroger sur l'avenir des normes qui seraient appliquées sur les métadonnées si

¹⁰ Pour une introduction aux concepts et techniques du Web Sémantique, lire l'article de BERNERS-LEE Tim, HENDLER James et LASSILA Ora, « The semantic web », in *Scientific American*, mai 2001.

les moteurs de recherche n'évoluent pas dans leur façon d'indexer et ne prennent pas plus en compte les métadonnées au détriment des textes intégraux qui sont actuellement leur support de travail. Par conséquent, il faudrait arriver à un consensus : d'un côté les normes doivent inclure de nouvelles méthodes comme créer des ontologies et de l'autre les moteurs doivent élaborer de nouveaux moyens de prendre en compte les balises. Une des solutions idéales serait de créer et d'implanter un langage du type thésaurus. Néanmoins, il semble a priori impossible de créer un tel langage qui conviendrait à l'universalité des acteurs, des sujets du Web. Pour l'instant on est loin de l'embryon même d'une véritable technique d'indexation des pages par les moteurs de recherche puisque ces derniers, pour la plupart, adoptent une indexation de type 0 ou 0+, comme défini par cette citation : « Il s'agit de l'*inversion* la plus sommaire qui soit : un lexique des mots du document (définis par les blancs et les ponctuations), est constitué en associant à chaque mot les adresses de ces occurrences dans le document. Bien qu'il n'y ait pas ici de traitements linguistiques à proprement parler, le niveau pris en compte de ce point de vue et celui du *découpage*. Il s'agit d'indexation libre. [...] Quelques variantes sont possibles à partir de ce type, nous les désignerons par « type 0+ » : il s'agit de l'ajout aux listes d'index des mots obtenus par changement de casse et suppression d'accents [...], ainsi que par phonétisation des index d'origine. »¹¹.

La manière de formuler les requêtes sur les moteurs de recherche est une difficulté à l'imposition d'une normalisation puisque elles sont en général effectuées par des non spécialistes et sont donc exprimées en langage naturel qui ne correspond pas forcément à un langage documentaire. Concrètement, la requête risque d'être incomprise car les termes employés ne correspondraient pas au langage imposé par les ontologies. Une solution serait que les concepteurs de moteurs de recherche conçoivent un « langage-passerelle » qui mènerait le langage naturel vers celui lié aux ontologies.

Enfin, les normes s'appliquent aux métabalises et très peu à leur contenu qui reste soumis à la logique de compétitivité qui régit l'Internet. Il serait donc intéressant d'approfondir l'étude des métadonnées afin de discerner une méthodologie qui permettrait à la fois d'être compétitif et de respecter les enjeux documentaires.

3. Le choix des mots-clés et les techniques d'indexation

On voit qu'en théorie, il existe de nombreux préceptes pour aider à l'indexation des pages Web. Mais il ne faut pas tomber dans l'excès inverse en manquant d'une certaine souplesse vis-à-vis de ces préceptes et donc en se fermant des portes. Une trop grande rigidité serait défavorable à la venue du public car le site serait alors sans doute moins bien référencé que si on effectue cette opération dans le cadre pratique duquel elle ne devrait pas être dissociée.

Un bon référencement de son site se prépare. Il faut tout d'abord bien définir les critères qui vont influencer sur le référencement.

¹¹ LEFEVRE Philippe, *La recherche d'informations : du texte intégral au thésaurus*, Paris : Hermès Science Publications, 2000, p. 109-110.

a) Préparer un site au référencement

La première chose à déterminer est les objectifs que l'on souhaite atteindre avec la création du site. Cela permet de se centrer sur les attentes principales et évite de se disperser à mettre au point des applications qui s'avéreront peu utiles et utilisées. C'est pourquoi, les objectifs doivent être fixés lors de la conception, ils doivent même aider à trouver une structure adéquate et un ton pertinent au site.

Le principal objectif et le premier qui vienne à l'esprit est de faire connaître l'instance pour laquelle le site est créé. Mais ça n'est pas toujours le cas. Le site Lille3 Jeunesse en est un bon exemple. Il n'était pas initialement destiné à être consulté par un public somme toute assez large. Cela montre bien qu'il faut évaluer fréquemment les objectifs remplis par le site puisque ces derniers peuvent évoluer. Un site conçu de manière confidentielle (et donc peu soucieux de sa place sur le Web) peut s'ouvrir au public et requérir un référencement plus poussé. L'adéquation des objectifs du site et de son public peut être vérifiée grâce à des outils statistiques qui renseignent le webmaster sur le nombre de visiteurs, leur provenance, la durée des visites, leur fréquence, etc. Une vérification régulière permet un réajustement rapide et moins fastidieux que si on ne fait pas attention à l'orientation que prend un site.

A l'instar de l'objectif, la cible doit être définie avant la création du site. Il est préférable d'établir une liste d'individus susceptibles d'être intéressés par le contenu que l'on va mettre en ligne. Garder un certain flou quant à la détermination du public peut paraître stratégique au premier abord, cela se révèle le plus souvent un faux calcul. Il est souvent plus porteur pour un site de répondre de manière précise et pertinente à une population spécifique. Circonscrire géographiquement la cible peut être utile pour certaines orientations à donner au site. Prenons l'exemple du site de l'office de tourisme de Lille. Si l'on sait que la population touristique qui visite cette ville est essentiellement hollandaise on pourra alors proposer dans la conception de traduire tout ou une partie des pages du site en néerlandais pour répondre à une demande de la part de ces utilisateurs.

Parallèlement à l'évaluation de l'objectif, il faut prendre en compte régulièrement l'évolution possible de la cible afin de recentrer le référencement. Il est évident qu'on n'indexera pas de la même façon un site destiné avant tout à des universitaires et un site dont le public visé est plus large.

La conception comporte une autre étape prépondérante. Elle doit se faire conjointement à une analyse de la concurrence. Cette dernière peut en outre être une source d'inspiration très intéressante pour son propre site. Il s'agit d'être critique et d'effectuer des recherches sur les moteurs de recherche et les annuaires les plus connus. Afin de partir sur les mêmes bases, il faut employer les mêmes mots-clés que ceux que l'on utilisera avec les moteurs de recherche, ou suivre un cheminement catégoriel identique à celui de notre site avec les annuaires. L'étude n'est pas obligatoirement exhaustive et peut très bien se baser sur par exemple les dix sites les plus pertinents. La régularité de ces recherches permet de suivre l'évolution des concurrents dans le domaine que l'on couvre et de vérifier par là même que l'on n'est pas dans une logique obsolète par rapport au sujet.

b) Méthodes pour le choix et l'insertion des mots-clés

(1) Le choix des mots-clés

Il existe plusieurs méthodes pour optimiser le choix des mots-clés qui seront utilisés pour le référencement d'un site.

On constate que beaucoup de développeurs ne se préoccupent de l'optimisation des moteurs de recherche qu'une fois la création du site achevée. Il ne semble pas que ce soit la meilleure approche : il vaut mieux construire un site autour de plusieurs mots-clés et d'un nom de domaine plutôt que d'attribuer ces derniers à un site préexistant. Une attention particulière doit être portée au répertoire et aux fichiers placés au sommet de l'arborescence du site puisqu'ils se verront accorder la pertinence la plus élevée par les moteurs de recherche. Ceux-ci vont référencer le site grâce au texte et aux mots-clés qui sont contenus dans ces fichiers. On peut donc dire que les mots-clés sont la pierre angulaire du référencement. L'opération de référencement consiste avant tout à indiquer aux outils de recherche les mots-clés sur lesquels le site souhaite être indexé.

La première étape pour attribuer des mots-clés est d'élaborer une liste réfléchie. Celle-ci est l'aboutissement d'un travail de recoupement de plusieurs listes.

La première liste est élaborée à partir du texte de la page que l'on souhaite indexer, il s'agit de dégager les dix mots qui apparaissent le plus dans le corps du texte. Un mot-clé proposé se verra attribué un meilleur score par les moteurs de recherche s'il est retrouvable dans le texte. Vient ensuite la liste éditeur pour laquelle on demande au professionnel gérant le site de proposer dix mots employés dans le métier et qui correspondent bien sûr au site. Une liste intuitive viendra en complément, on se met à la place de l'utilisateur et on retient les mots que l'on croit susceptibles d'être utilisés dans ses requêtes. Pour cette liste, il ne faut pas prendre des termes trop génériques et il faut essayer de toujours raisonner en associations. C'est le moment idéal pour inclure des termes à l'orthographe erronée (mais qui pourrait être fréquemment employée si le terme correctement orthographié est assez rare et difficile à écrire), aux synonymes et aux termes en rapport lointain (ce qui peut s'avérer une tactique dangereuse si le moteur de recherche ne perçoit aucun rapport avec les thèmes du site). Il est toujours intéressant de chercher une source d'inspiration chez les concurrents en choisissant en priorité ceux qui apparaissent en meilleure position aux requêtes formulées avec les mots-clés choisis a priori pour son propre site. L'étape suivante consiste à mettre les listes en commun, on enlève les termes trop génériques et on garde de dix à vingt mots. La société Worldtracker propose sur leur site un outil, le Search Term Suggestion Tool, qui permet de tester l'efficacité d'un mot-clé en calculant son Keyword Effectiveness Index (KEI), en français, « indice d'efficacité d'un mot-clé ». C'est le rapport entre le nombre de fois où un terme apparaît dans la base de données de Worldtracker et le nombre de pages qui y font référence.

Il est préférable d'associer des mots-clés que de mettre un seul mot-clé, c'est le regroupement par proximité. Cette règle pourra souffrir quelques exceptions si le site se base sur un néologisme, un acronyme, une marque connue. Cette façon de faire résulte d'une tendance actuelle de combiner plusieurs mots pour formuler une requête. Cependant, il ne faut pas tomber dans l'excès inverse et il vaut mieux se limiter à deux ou trois phrases clés afin de conserver une pertinence accrue et des réponses affinées aux requêtes des utilisateurs.

Une fois la liste achevée il convient de tester les mots-clés qu'elle contient à l'aide de requêtes formulées avec eux sur les moteurs de recherche. Le test doit porter sur toutes les combinaisons possibles (en solo, en duo, en groupe). On enregistre le nombre de pages ressorties en résultat pour chaque configuration. Les meilleurs mots-clés sont ceux qui

permettent l'affichage d'un nombre moyen de pages. En effet, trop de pages signifie trop de bruits, trop peu de pages, trop de silence. On peut encore affiner le test en recoupant les résultats obtenus par un moteur avec un autre. On appelle cela le « test de résonance ». A la fin, on désigne cinq mots-clés primaires et cinq secondaires. Cela se révèle suffisant puisque les moteurs prennent essentiellement ceux-ci en considération. Les mots restants peuvent être placés de manière complémentaire.

(2) *L'insertion des mots-clés*

Les moteurs de recherche suivent un ordre bien précis dans leur façon de répertorier les pages Web. Ils se basent avant tout sur la manière dont sont utilisés les mots-clés dans une page.

Il existe un ordre de priorité dans le classement des mots-clés dans une page Web par un moteur de recherche : le nom de domaine (ou adresse Internet), la balise <TITLE>, les balises <H1> à <H6>, le contenu de la page, les métabalises, les liens proposés. L'ordre suivi pour classer les sites est le suivant : la prééminence, la fréquence, la densité, la proximité, l'emplacement, ...

Une fois les mots-clés choisis et ordonnés, on commence à les disposer dans le code de la page. Les mots-clés primaires sont placés en priorité dans la balise <Title> à laquelle le moteur donne le plus de poids. Les autres mots-clés sont placés dans la métabalise *Keywords* (dans le code, elle se note <META NAME = « keywords » CONTENT = « mot-clé1, ... »>) et enfin dans la métabalise *Description* (dans le code, elle se note <META NAME = « description » CONTENT = « blabla »>). Il faut faire attention et employer les virgules avec parcimonie. Celles-ci ne s'imposent vraiment que lorsque deux termes désignent des concepts différents. Il faut être vigilant à la lemmatisation. Il se révèle souvent plus efficace de préférer la forme au pluriel à la forme au singulier d'un mot puisque dans la majorité des cas la chaîne de caractères du dernier se retrouve dans le premier uniquement complétée par la désinence du pluriel. Par exemple le terme « chaussure » ressortira systématiquement lors d'une requête même si « chaussures » est indexé mais l'inverse, c'est-à-dire une requête « chaussures » alors que l'indexation a été faite avec « chaussure » ne sera pas donné en résultat car l'ordinateur n'aura pas établi de correspondance entre les deux chaînes de caractères. Ce précepte s'applique également pour le féminin. Dans le cas où le mot masculin est contenu dans le féminin, il vaut mieux indexer avec ce dernier pour qu'une requête formulée avec n'importe lequel des deux genres aboutisse à un résultat. Dans le cas où le terme au pluriel ou au féminin est très différent du singulier ou du masculin, il vaut mieux faire coexister les deux termes dans la liste des mots-clés. Cette technique est appelée *stemming*.

Il existe un avantage indéniable d'inclure dans le code de la page une métabalise *Description*. Celle-ci ne doit pas excéder 200 caractères (soit environ 25 mots) puisque au-delà de cette limite ils n'apparaissent pas sur les écrans de résultat et ils sont de toute façon moins pris en compte par les moteurs. Cependant, cette balise reste très importante car elle permet de faire apparaître un résumé explicatif du site sur l'écran de résultat d'une requête. La métabalise *Description* ne doit pas répéter le titre et être si possible différente pour chaque page d'introduction d'une partie importante du site.

Mais l'insertion de métatags se révèle insuffisante pour obtenir un bon positionnement. Elles sont en effet de moins en moins prises en compte suite à leur utilisation intempestive. C'est pourquoi, une fois les métabalises renseignées, il n'est pas interdit et il même fortement conseillé de compléter l'indexation de son site en plaçant certains mots-clés à des endroits stratégiques de la page. Les mots-clés ne se placent pas de manière aléatoire dans le code d'une page. Il existe certaines règles à connaître pour optimiser l'indexation de

son site. Les moteurs de recherche prennent en effet plus en considération certaines zones que d'autres pour indexer les pages. La zone qui a le plus de poids est celle du titre (après le nom de domaine) puisque les moteurs de recherche lui octroient une grande valeur lors de leurs calculs (il est important de préciser que la balise <TITLE> n'est pas une métabalise). Un titre doit être descriptif, fonctionnel et concis (environ cinq mots-clés qui s'associent parfaitement). Cette zone est indexée dans sa totalité par les moteurs de recherche et les annuaires mais seuls 55 à 90 caractères (70 à 80 en moyenne) sont affichés dans les résultats de requêtes.

Le classement des pages Web dans l'ordre alphabétique et en particulier les titres de ces pages peuvent être influencés par la valeur ASCII des caractères¹². En effet, seuls les moteurs de recherche francophones prennent l'accentuation des caractères en compte (ces lettres accentuées sont alors ramenées à leur valeur sans accent). Il existe donc un risque de voir un mot-clé comportant un accent non pris en compte par un moteur de recherche anglophone. De même les espaces, les caractères spéciaux puis les chiffres ont une valeur moindre par rapport aux lettres. On peut donc essayer de biaiser les moteurs de recherche en insérant un de ces signes typographiques avant un titre. Mais cela peut se révéler dangereux puisque les moteurs sont accoutumés à ces procédés frauduleux et peuvent décider d'un bannissement de la page ou du site s'ils en repèrent l'emploi. Toutefois, tous les moteurs de recherche ne prennent pas en compte l'ordre ASCII.

La balise <!--commentaire--> (destinée comme son nom l'indique à recueillir des commentaires) est initialement prévue pour recevoir des annotations quelconques et qui n'apparaissent pas lors de la visualisation. On peut détourner quelque peu cette utilisation et y inscrire des mots-clés. Ce procédé ne fonctionne pas avec tous les moteurs de recherche mais il donne la possibilité d'augmenter la densité des mots-clés sans pour autant les faire apparaître dans la page. Il présente en plus l'avantage de permettre le référencement de sites programmés en Flash.

Les images peuvent également constituer une source de contenu à indexer. En effet, actuellement, les moteurs de recherche ne tiennent pas compte de leur contenu visuel. Il faut donc les décrire, ce qui se fait avec la fonction ALT dans la balise . Cela permet en outre à l'utilisateur de voir la description d'une image avant que celle-ci n'apparaisse lors du chargement d'une page. Dans le cas d'images réactives, il faut ajouter des liens textes en dessous en plus des fonctions ALT afin de permettre aux robots et aux internautes qui désactivent ces images de suivre les liens qu'elles contiennent. Les *imagemaps* sont en général acceptées par les moteurs à condition de comporter un minimum d'informations dans le texte accompagnant sous peine d'être refusées par les moteurs de recherche.

Une dernière méthode consiste à inclure des mots-clés dans le texte de la page en gardant à l'esprit que des mots mis en liens, en gras ou mis en forme avec une balise de type <H1> seront privilégiés, surtout s'ils sont en haut de page. Idéalement, une page moyenne possède entre 400 et 700 mots. Il faut veiller à ce que la densité des mots-clés soit comprise entre 3 et 10% sous peine d'être bannis par le moteur. De plus, il est fortement déconseillé d'avoir recours au *spamdexing* qui lui aussi entraîne le bannissement mais on conseille de répéter les mots servant à l'indexation dans toutes les parties pour mettre en avant les mots-clés forts, sans toutefois dépasser 4 ou 5 répétitions par page.

On peut aussi multiplier les accès au site dont on s'occupe en créant des pages « passerelles » ou pages « alias » qui sont des duplicata modifiés de manière minime d'une page existante sur laquelle on souhaite faire aboutir un visiteur. Ce procédé a un but bien précis. L'exemple le plus courant est de proposer une version multilingue d'un même site ou d'une même page. Cela permet de multiplier le référencement d'un même site vers des cibles

¹² Voir Annexe 7 : Valeur ASCII des caractères.

différentes, d'optimiser le référencement vers des outils de recherche ciblés et de revaloriser le positionnement des pages sur les outils.

Les noms de fichiers étant également référencés il est préférable de les rendre intelligibles.

On peut également acheter plusieurs noms de domaine en rapport avec l'activité du site si c'est faisable.

(3) Les robots

Certaines pages, comme les pages privées, les pages confidentielles, les pages avec des tarifs, etc. n'ont aucun intérêt à être soumises à l'indexation. Or les robots envoyés par les moteurs de recherche procèdent à l'indexation de toutes les pages et au suivi de tous les liens. Le classement des sites peut, de ce fait, se trouver pénalisé, et les scores attribués peuvent être influencés. En conséquence, il peut s'avérer judicieux d'exclure de telles pages du référencement.

Pour ce faire, il existe deux méthodes qui visent à contrôler la visite du robot sur un site. On peut tout d'abord utiliser des métabalisés inscrites comme suit dans le code de la page : <META NAME « Robot » CONTENT = « contenu »>. Il faut remplacer « contenu » par [NO]INDEX pour indexer ou non la page et/ou [NO]FOLLOW pour suivre ou non les liens de la page. L'autre méthode, plus précise au niveau du contrôle, consiste à inclure un fichier nommé robots.txt à la racine du serveur. Cette méthode est régie par la norme RES (Robot Exclusion Standard) créée en 1994 pour indiquer les répertoires à ignorer. Un fichier robots.txt se compose comme suit :

```
User-agent : *  
Disallow : /répertoire/  
Disallow : /répertoire/sous-répertoire/
```

« User-agent » indique le robot cible (si tous les robots sont concernés, on ajoute le signe « * », sinon il faut stipuler le nom du robot désigné). « Disallow » indique les chemins à ne pas suivre, pour tout interdire, il faut écrire « Disallow : / ». Inversement, si l'on veut tout accepter on utilisera « Disallow : ». Mettre un « / » en fin de ligne autorise l'indexation des fichiers du répertoire désigné, mais pas de ses sous répertoires. A l'inverse, ne pas mettre un « / » interdit l'indexation du répertoire en entier.

Un autre moyen d'interdire l'accès d'une page à un robot est d'y ajouter un mot de passe.

(4) Les techniques de fraude

On constate donc qu'il existe plusieurs moyens d'indexer. Certains restent plus ou moins tolérés, au webmaster de rester vigilant à ne pas pousser trop loin la limite et se faire bannir. Par contre, il y a des procédés clairement proscrits exploitant les failles des moteurs de recherche. S'ils sont utilisés, ils entraînent une sanction inévitable dans le cas où le moteur les repère. La peine varie en fonction de la gravité de la tentative de la fraude et des moteurs puisqu'ils ont chacun leur propre politique.

La technique de fraude la plus courante est le *spamdexing*. On peut aussi répéter la balise <TITLE> qui est celle que le moteur prend le plus en considération, incruster du texte invisible ou avoir recours au « JavaScript retour ». Cette méthode consiste à rediriger automatiquement les robots vers une « page passerelle » (ou alias) avec suffisamment de contenu pour que la fraude ne soit pas perceptible. Le pixel invisible fait également partie des « arnaques » : on insère dans la page une image d'un pixel de côté et de la même couleur que le fond associée à de nombreux mots-clés avec la fonction ALT.

Une technique très poussée est « l'amorce et l'échange ». Une page indexée efficacement sera immédiatement modifiée et remplacée par une autre tout en conservant la même adresse une fois référencée par un moteur. Une dérive existe, il s'agit de « l'échange-serveur » ou *cloacking*. Cela fonctionne sur le même principe que la fraude précédente. Un robot visitant une page est repérable par son adresse IP. Le site frauduleux qui a identifié le robot va le renvoyer vers une autre page entièrement configurée pour répondre à toutes les attentes de celui-ci et ainsi obtenir un bon score. Il peut être fait de même pour chaque robot en fonction des exigences particulières des moteurs de recherche.

Il y a encore le formulaire caché, avec lequel on peut rajouter une liste de mots-clés en insérant une balise <INPUT>. Enfin, les liens cachés sont un dernier moyen de contourner l'intégrité du moteur.

Mais quelles sont les sanctions encourues ? Un élément reconnu comme du *spamdexing* n'entraîne pas forcément la non indexation de la page qui le contient mais celle-ci se verra inéluctablement pénalisée dans le classement. Dans les cas les plus abusifs, elle peut purement et simplement ne pas être indexée, voire entraîner la suppression de l'indexation d'autres pages enregistrées avec un nom de domaine identique. Si la fraude est vraiment très importante, le nom de domaine peut même se retrouver sur une « liste noire » qui empêche toute nouvelle tentative de référencement jusqu'à révision du jugement par le moteur. Un renouvellement d'inscription pour une même page proposé trop rapidement peut également entraîner des sanctions. Enfin, un dernier comportement délictueux peut poser problème, il s'agit du recopiage de code de sites populaires ou renommés qui connaissent un bon classement.

Il est par contre très difficile, voire impossible, de tricher avec les annuaires qui sont gérés manuellement par des documentalistes. Ils n'effectuent pas leur classement d'après le code de la page Web mais d'après la qualité de son contenu. Toute modification du code ne sera donc pas prise en compte.

B. Le référencement

Référencer son site consiste à faire la démarche de le proposer à un moteur de recherche ou à un annuaire pour qu'il l'intègre dans sa base de données.

1. Intérêt du référencement

« L'intérêt du référencement n'est pas de permettre à une entreprise d'être trouvée par les outils de recherche par son nom ou sa marque, mais plutôt de permettre à quelqu'un qui

recherche un service ou un produit de trouver rapidement cette entreprise alors qu'il ne la connaissait pas ou bien qu'il n'y pensait pas. »¹³

Face à l'engorgement du réseau, il convient de mettre au point une bonne méthode de référencement si l'on veut pouvoir être retrouvé. En effet, actuellement on assiste à un allongement des délais pour le référencement, les annuaires sont en manque de personnels pour gérer l'afflux des demandes. On assiste à une limitation du volume journalier des soumissions de manière drastique et à l'émergence de sous-traitants tels que Inktomi. Le ralentissement du rythme d'indexation, la surcharge due au retard pris par les moteurs de recherche dans leurs traitements et l'allongement des files d'attente nécessitent plus que jamais la mise en place d'une méthode organisée et systématique pour la soumission et l'entretien du positionnement.

Avec plus de trois milliards de pages sur Internet les outils de recherche sont indispensables aux internautes pour trouver les informations qui les intéressent. Mais parallèlement, cela veut aussi dire que la promotion d'un site Web doit se faire de manière plus appliquée pour avoir une chance de voir son site surnager par rapport à la masse. Il faut avant tout se soucier d'avoir un bon positionnement dans les moteurs de recherche plutôt qu'un trafic important puisque le premier générera automatiquement le second.

2. Aperçu des méthodes de recherche d'information sur le Web

La recherche sur l'Internet est étroitement liée aux moteurs de recherche et aux annuaires utilisés. Chacun possède en effet ses règles, ses syntaxes et ses références. Ils n'ont pas tous les mêmes ressources et les mêmes manières de les exploiter. En outre, les modes de recherche proposés à l'utilisateur varient énormément d'un outil à l'autre.

La majorité des requêtes formulées sur le Web n'est pas composée comme on pourrait s'y attendre par un petit nombre de requêtes ayant une très grande fréquence, mais plus par un grand nombre de requêtes rares (singulières). En d'autres termes les requêtes fréquemment posées ne constituent qu'un petit échantillon de la totalité des requêtes. Ces requêtes sont caractérisées par la singularité, c'est-à-dire que le mot utilisé ne se retrouve qu'une ou un très petit nombre de fois sur la base de données globale. Comme ce nombre de mots peu utilisés est gigantesque, on parle également de diversité pour qualifier les requêtes. Cela montre que ces dernières ne sont pas soumises à des effets de mode ou de publicité et qu'il faut se garder d'inclure dans sa politique de référencement un poids démesuré aux requêtes dites populaires qui ne sont en fait qu'une infime partie des requêtes faites sur le Web¹⁴.

Quatre logiques président aux recherches des outils : la recherche géographique avec des sites répertoriés par ville, la recherche thématique avec un catalogue thématique consulté à l'aide de mots-clés (comme Yahoo !), la recherche par index employée par les moteurs de recherche mais qui présente l'inconvénient de ne concerner que moins de 50% du contenu de l'Internet, et la recherche par méta-index. Utilisée par les méta-moteurs, elle combine les recherches parallèles de plusieurs moteurs de recherche et elle restitue un résultat sous forme compilée¹⁵.

¹³ ESKENAZI Jean-Pierre F., *Référencement : comment référencer son site web*, Versailles : Webedition, 1999, p. 48.

¹⁴ Selon J. LAJOIE, Richesse et diversité des requêtes, in *Comprendre les usages de l'Internet*, Paris : Éditions Rue d'Ulm, 2001, p. 191-194.

¹⁵ Selon SAMIER H., SANDOVAL V., *La recherche intelligente sur l'internet et l'intranet*, 2^e éd. revue et augm., Paris : Hermès Science Publications, 1999, p. 14-20.

Il a été vu que les recherches se font à l'aide d'outils de recherche qui peuvent se diviser en deux catégories principales : les annuaires et les moteurs de recherche.

Un annuaire se compose de rubriques et de sous-rubriques disposées par thèmes (on parle aussi d'annuaire thématique). Le mode de recherche tient plutôt du cheminement avec ce genre d'outils, il faut aller de rubrique en rubrique, du général au particulier. On aboutit sur une sélection de liens. Cependant, les annuaires proposent de plus en plus systématiquement un moteur de recherche qui agit sur leur base de données. L'annuaire ne fournit que très peu d'informations sur les sites qu'il met en lien (le nom, l'adresse, une description succincte, la catégorie). La particularité principale de cet outil est de proposer un traitement humain de l'inscription d'un site, ce qui a pour conséquence un nombre plus limité de sites à disposition dans leur base mais une qualité accrue des sites proposés. Exemples d'annuaires internationaux les plus populaires : Looksmart, Open Directory Project, Yahoo ! Des annuaires francophones : Nomade, Yahoo ! France.

L'autre grande catégorie d'outils de recherche est le moteur de recherche. Le traitement de l'information est ici totalement automatisé. Ce sont des robots qui enregistrent les URL, la description, les mots-clés et les textes des pages qu'ils trouvent en *surfant* de lien en lien. Les moteurs effectuent leur recherche sur l'intégralité du texte des millions de pages de leur base de données aussi appelée index de recherche. Celle-ci est de telle ampleur que sa mise à jour demande plusieurs semaines et voit les délais s'allonger. Les résultats d'une requête sont fournis sous forme de résumé des pages dont la pertinence est déterminée par une série de formules (algorithmes). Chaque moteur a son propre index et algorithme. Exemple de moteurs de recherche internationaux les plus populaires : Altavista, Excite, Fast-Alltheweb, Go.com (ancien Infoseek), Google, HotBot, NorthernLight, WebCrawler. Francophones : Altavista France, Ecila, Lokace, Lycos, Voila.

3. Le Web invisible

Malgré l'existence d'outils pour référencer les sites, une bonne partie de ceux-ci reste plus ou moins inaccessible à l'utilisateur, c'est ce qu'on appelle le Web invisible. Chaque moteur couvre au mieux 10% (pour les plus populaires) de la masse totale d'information du Web. Des études ont démontré que seuls 45% du Web environ sont indexés si l'on prend toutes les bases de données de tous les moteurs car de nombreux résultats se recoupent. Et les prévisions annoncent une diminution de ce taux. On peut distinguer deux sortes de Web invisible : la première se constituerait de pages inaccessibles comme les pages solitaires (sans liens entrant ou sortant), celles qu'on refuse de référencer (en bloquant l'accès aux robots), les intranet protégés, etc. La seconde est définitivement inaccessible : les bases de données consultables via des serveurs ou le Web dynamique programmé avec des langages que les moteurs ne comprennent pas encore (pages en Flash, ...).

Cette idée de Web inaccessible a longtemps été éclipsée par l'idée prédominante que le Web avait une forme de toile dans laquelle toute page pouvait se retrouver à partir d'une autre (en 19 clics paraît-il). C'est en 1999 que des recherches prouvent le contraire et met au jour une structure en « nœud papillon »¹⁶. Les pages étant identifiées par les moteurs en fonction uniquement des liens qu'elles comportent, il reste de nombreuses pages déconnectées que les moteurs ne peuvent indexer. D'autres part, les délais de renouvellement des visites des robots sur un site sont de plus en plus espacés, ce qui risque de faire oublier certaines pages créées ou modifiées après l'un des passages. Ce cheminement de lien en lien n'est pas

¹⁶ Voir Annexe 10 : Structure du Web en « nœud papillon ».

impartial : il se fait à l'avantage des sites les plus connus et au détriment des sites récents ou spécialisés moins susceptibles d'être reliés aux autres¹⁷.

Un début de réponse est apporté avec les annuaires sélectifs qui sélectionnent les sites indexés le plus souvent par des professionnels de l'information (bibliothécaires, documentalistes). Ces sites sont essentiellement à caractère professionnel et doivent répondre aux questions des universitaires et des enseignants. Chaque site fait l'objet d'une évaluation qualitative et est décrit de manière assez poussée (on peut stipuler les rubriques proposées, les possibilités de recherche, ...). Toutefois, les annuaires sélectifs sont très peu en français. L'exemple le plus connu est celui des « Signets de la BNF »¹⁸.

Il existe aussi des annuaires spécifiques, c'est-à-dire des annuaires spécialisés dans un seul domaine. De plus en plus populaires, ils permettent une approche complète du sujet et évitent une perte de temps lors des recherches en regroupant tous les sites. Il peut donc s'avérer judicieux de s'inscrire sur ce type d'annuaires. On peut les retrouver grâce à deux annuaires d'annuaires spécifiques, l'un anglophone (www.beaucoup.com), l'autre francophone (www.enfin.com).

4. Méthodes pour le référencement d'un site Internet

Il a été vu combien l'opération de référencement est importante pour assurer la fréquentation de son site. L'inscription sur les outils de recherche s'inscrit dans une politique de marketing. Il s'agit d'un marketing « semi-direct », le marketing « direct » étant de faire de la publicité en dehors de l'Internet (carte de visite, publicité dans les magazines spécialisés, etc.).

Il faut compter avec de nouveaux critères lors du référencement. Si à l'origine seul le contenu visible d'une page était pris en compte, les abus constatés ont engendré l'émergence de nouvelles caractéristiques indépendantes du contenu avec lesquelles il est plus difficile de tricher (la popularité, le comportement de recherche des internautes (indice de clic), la présence ou non du site dans les annuaires des moteurs de recherche, ...). Parallèlement, on a vu le développement de critères de rejet tels que l'utilisation de *frames* (cadres) sans la balise <NOFRAMES>, l'utilisation de scripts CGI sur la page d'accueil, l'utilisation d'une page trop graphique (c'est-à-dire de moins de 50 mots, en Flash par exemple), l'utilisation de la balise <METAREFRESH>, l'utilisation d'un JavaScript de redirection ou la redirection du nom de domaine, l'utilisation d'une URL compliquée avec des caractères comme ?, \$, &, %, § (souvent pour les sites dynamiques). A la vue du nombre de critères pris en compte, il semble plus judicieux d'évaluer les pages qu'un webmaster choisit pour être « candidates » au référencement.

Il convient de tenir compte d'une certaine contrainte de temps : proposer l'inscription d'une page de son site à un outil de recherche ne signifie pas que celui-ci d'une part va automatiquement le prendre en compte et d'autre part qu'il le fera dans les délais les plus brefs. L'attente peut varier de 24 à 48 heures pour les plus rapides jusque 8 à 10 semaines (et plus) pour les plus lents. Il faut donc tabler sur une période de trois mois pour effectuer un référencement efficace.

¹⁷ FOENIX-RIOU Béatrice, *Recherche et veille sur le Web visible et invisible : agents intelligents, annuaires sélectifs, interfaces des grands serveurs, portails thématiques*, Paris : Bases, Paris : Éditions technique et documentation, 2001, p.11.

¹⁸ FOENIX-RIOU Béatrice, *Recherche et veille sur le Web visible et invisible : agents intelligents, annuaires sélectifs, interfaces des grands serveurs, portails thématiques*, Paris : Bases, Paris : Éditions technique et documentation, 2001, p.11.

Se pose ensuite le problème du choix d'un traitement manuel ou automatique, ce qui revient à penser en termes respectifs de qualité ou de quantité. L'inscription automatique propose le référencement auprès de 600 à 1500 moteurs de recherche. Cependant, la majorité des moteurs choisis sont de petite envergure au point que leurs listes de liens ont un trafic plus dense au niveau des inscriptions qu'au niveau des consultations. L'avantage est de pouvoir augmenter son trafic et donc sa popularité grâce à l'inscription sur des petits moteurs mais le principal inconvénient est de n'avoir aucune certitude quant à la prise en compte de sa demande d'inscription (car de plus en plus de moteurs repèrent ces requêtes automatiques et les rejettent systématiquement) et aucun contrôle sur les informations soumises. De plus, ils ne gèrent pas les inscriptions auprès des annuaires et ne sont pas efficaces avec des moteurs non anglo-saxons. Ces outils d'inscription automatique se présentent sous forme de logiciel ou de site spécialisé dans ce service. La plupart sont payants. Le logiciel SubmitWolf est le leader avec la proposition d'inscription chez 1700 moteurs de recherche. Il offre beaucoup de paramètres. Sa version francophone s'appelle ProMoteur. Webposition Gold propose peu de moteurs mais offre de nombreux services comme la création de pages passerelles, une analyse du positionnement, du code HTML, les statistiques du trafic, ...

On peut s'inscrire « manuellement » aux différents outils de recherche. C'est effectivement la seule démarche efficace pour un référencement optimal. Une opération d'inscription demande de la préparation, il peut se révéler judicieux de mettre de côté toutes les informations nécessaires : le titre du site, les URL (celle de la page d'accueil et des pages principales du site), un e-mail pour pouvoir joindre un responsable du site, la description du site, les trente principaux mots-clés, le nom des principaux concurrents pour pouvoir repérer plus facilement la catégorie correspondant à son site dans les annuaires, la date de lancement du site.

Deux méthodes peuvent être appliquées, une complète et une simplifiée par rapport à la précédente. La première consiste à inscrire sa page d'accueil ainsi que les principales pages du site aux principaux moteurs de recherche et annuaires internationaux et à des outils choisis préalablement selon un critère géographique en fonction de la zone ciblée. On peut ensuite utiliser les outils de référencement automatique pour proposer la page d'accueil uniquement à tous les autres outils de recherche possibles. Il faut patienter environ six semaines, le temps que toutes les mises à jour soient faites et on peut vérifier les inscriptions en effectuant des recherches sur les moteurs sélectionnés en utilisant les mots-clés inclus dans la page. Il faut procéder à une réinscription dans les outils de recherche qui n'ont rien pris en compte et modifier les pages qui ont posé problème, quitte à créer des pages passerelles. On peut également mettre ces six semaines à profit et analyser la fréquentation du site afin de tenter de dégager quels sont les mots-clés qui fonctionnent le mieux et les sites qui nous amènent le plus de public. La version simplifiée ne demande d'inscrire que la page d'accueil aux principaux moteurs et annuaires internationaux et ceux ciblés selon un critère géographique. On peut patienter six mois et relancer les tentatives d'inscription dans les outils qui n'ont rien pris. Cette méthode peut être employée surtout si on n'est pas pressé par des indices de popularité¹⁹.

Il existe plusieurs raisons à la non prise en compte de sa page. Il peut y avoir une surcharge de volume des demandes à traiter par le moteur de recherche (dans ces cas-là l'inscription peut simplement être ignorée), des conflits de paramétrage. Si des efforts répétés n'aboutissent à rien, il faut écrire au webmaster de l'outil pour tenter de savoir si on est en accord avec les règles de l'outil.

Il y a quelques écueils à éviter et quelques conseils de base à suivre. Il ne faut jamais inscrire un site en construction. Il vaut mieux établir un tableau à double entrée (une « outils »

¹⁹ Voir Annexe 8 : Schéma d'inscription pour une page sur un moteur.

et une « pages référencées ») et indiquer les dates d'inscription effective dans les cases pour mieux suivre l'opération de référencement. En effet, les sites qui soumettent deux pages en moins de 24 heures peuvent être bannis. Si l'on désire être inscrit sur un outil de recherche étranger, et non pas international, comme Yahoo ! Allemagne, il faut quasiment obligatoirement proposer une version du site dans la langue du pays et posséder un nom de domaine dont l'extension correspond à celle du pays (www.monsite.de dans l'exemple de Yahoo ! Allemagne). Lors de la suppression d'une page indexée, il faut penser à déréférencer cette page auparavant ou à créer une page de transition vers la nouvelle, autrement l'ancienne page subsistera dans la base de données du moteur le temps de la mise à jour du robot. Celui-ci rejetant systématiquement les pages qui ne mènent à rien (erreurs 404), il faut procéder au référencement de sa page pour la faire aboutir à rien si l'on veut avoir une politique de déréférencement efficace.

5. Les obstacles au référencement

Des difficultés persistent durant l'opération de référencement, dues aux outils de recherche qui ne sont pas parfaits. Elles surviennent inmanquablement quand on propose un site un peu plus élaboré.

Les problèmes rencontrés avec la lisibilité du sens des images a été abordé précédemment. Les robots actuels ne savent toujours pas comprendre une image et donc l'indexer. C'est un des objectifs pour les futurs moteurs.

Les *frames* (ou cadres en français) ne sont supportés que de manière partielle par les moteurs de recherche qui ignorent toutes les informations contenues entre les balises <FRAME> et <FRAMESET>. Une solution consiste à s'efforcer de limiter ce qui sera lu par le robot dans la *frameset* (c'est la page-ressource du contenu des fenêtres de la page). Il faut veiller à y intégrer les informations concernant le titre, un texte chargé de mots-clés et les liens vers toutes les autres pages du site entre les balises <NOFRAMES>. De la sorte le robot venu visiter la page pourra en référencer le contenu et suivre les liens qui y sont précisés pour référencer les autres pages auxquelles ils renvoient²⁰.

Les pages dynamiques posent également problèmes. Elles sont généralement générées par des bases de données (ou des langages exploitant ce fonctionnement). Elles utilisent fréquemment des *applets* Java. Les pages dynamiques sont réalisées à l'aide de la norme CGI (Common Gateway Interface). Or les moteurs de recherche rejettent les URL contenant les chaînes de caractères *cgi-bin*, \$, ?, % ou d'autres symboles car ces derniers ne renvoient pas à des informations en HTML et donnent en outre accès à la base de données intégrée au site, trop volumineuse en terme d'informations contenues pour qu'un robot prenne le loisir de s'y perdre. Une des solutions peut être de créer une page alias statique que l'on soumettra à l'inscription.

6. D'autres techniques de promotion d'un site Internet

Il est possible de promouvoir un site autrement que par les outils de recherche et le référencement.

²⁰ Voir Annexe 9 : Code d'une page en HTML corrigée pour gérer au mieux les *frames*.

On sait que la popularité est un critère de plus en plus pris en compte pour déterminer le classement d'une page. Ce qui veut dire qu'un site récent a peu de chance de voir de nombreux liens pointer vers lui et d'augmenter ainsi sa popularité.

Il existe plusieurs méthodes de marketing indirect pour pallier cet inconvénient.

Les échanges de liens constituent la meilleure des promotions. Il s'agit d'un échange de bons procédés : un site fait pointer un lien vers le nôtre, en échange, le webmaster s'engage à appliquer la réciproque. Le but est d'avoir un retour important d'utilisateurs. Leurs visites sont d'une grande qualité puisque l'internaute qui accède au site par un lien est susceptible de porter un intérêt particulier au contenu de ses pages. Le site gagne en crédibilité à être cité par les autres et son score augmente d'autant plus qu'il existe de liens pointant vers lui. La popularité s'en trouve accrue, ce qui peut affecter le classement dans certains moteurs. Il faut veiller à créer une page consacrée aux liens pour accueillir ceux obtenus par échange quitte à l'enrichir avec des liens vers des outils de recherche. Cette page doit être placée en deuxième ou troisième niveau de l'arborescence et surtout pas sur la page d'accueil et elle doit contenir entre 20 et 100 liens pour amorcer l'efficacité de la démarche. Il faut également faire attention à générer l'ouverture d'une fenêtre lorsque l'internaute clique sur le lien pour qu'il ne quitte pas le site. Il faut vérifier régulièrement la qualité des liens inclus dans le site. En effet, des liens obsolètes (erreurs 404) produiront forcément un mauvais effet sur les utilisateurs qui pourront s'interroger à juste titre sur la validité de la mise à jour du site. Cette vérification doit s'étendre à la réciprocité à laquelle s'engagent les sites partenaires. Il n'y a en effet dans ce cas là aucune raison d'héberger un lien vers un site qui ne rendrait pas la pareille. Pour trouver des partenaires, il faut aller voir chez les concurrents et leur proposer à eux et aux sites ayant des liens avec eux l'échange de liens. Dans le cas d'une page Web pointant vers son site mais qui serait mal ou pas référencée par les outils de recherche on peut même proposer la réinscription de cette page à l'insu de son webmaster afin d'augmenter indirectement la popularité de son site.

Un autre bon outil de promotion est le *Webring* grâce auquel on peut gagner sans efforts de nouveaux visiteurs. Il s'agit d'une alliance d'au moins cinq sites traitant d'un même sujet qui se groupent entre eux pour offrir la possibilité de naviguer de l'un à l'autre, parfois aléatoirement, sans avoir à formuler une requête. La navigation est rendue plus aisée et les utilisateurs sont satisfaits de ne pas avoir à trier les résultats proposés par les moteurs de recherche. Cela explique l'expansion de ce service.

Il est aussi possible de passer par les *newsletters* (ou *e-zine* ou encore liste de diffusion). Celles-ci peuvent adopter la forme d'un site Internet ou de l'envoi d'e-mails (ou encore les deux à la fois). Il faut néanmoins trouver la *newsletter* qui correspond au thème de son site pour que ce procédé soit efficace. Le principe réside dans une mutualisation de petites annonces que tous les membres peuvent écrire et consulter. Cependant, comme aucun but commercial ou lucratif ne doit motiver l'adhésion à une *newsletter*, il faut rédiger un article susceptible d'intéresser un grand nombre d'internautes dans lequel on laissera les coordonnées de son site. Une autre solution peut être de créer sa propre *newsletter* afin de fidéliser les visiteurs qui le souhaitent.

Un autre moyen d'assurer sa promotion est d'utiliser les conférences de presse virtuelles. Certains sites font office d'intermédiaires entre son site et la presse spécialisée et généraliste auxquelles ils envoient des e-mails pour présenter le site qui a sollicité leurs services.

On peut aussi se lancer dans une campagne de publicité avec les bannières. Cela consiste à acheter un espace publicitaire, la bannière, qui fera la promotion d'autres sites ; en contrepartie, son site peut apparaître dans la bannière incrustée dans la page d'un autre site. Cependant, cette méthode se révèle coûteuse et aléatoire, surtout pour des sites de petite envergure puisque la publicité se fait proportionnellement à l'ampleur du site.

7. Comment maintenir le niveau de popularité

La popularité une fois acquise n'est pas pour autant pérenne si on ne fait rien pour l'entretenir. Des milliers de sites se créent chaque jour, ils peuvent détrôner à chaque instant un site pourtant bien classé dans les outils de recherche. Rien n'est immuable, les bases de données des outils de recherche se modifient constamment, il en est de même pour les règles d'indexation qui voient leurs algorithmes sans cesse changer en vu d'une optimisation. Il n'est pas étonnant que le classement soit affecté et que le positionnement évolue. C'est pourquoi, il faut instaurer une veille permanente et vérifier très régulièrement sa position. Si celle-ci diminue, il faut reproposer les mêmes pages Web à l'inscription. Il faut aussi savoir s'adapter en créant de nouvelles pages alias dont les mots-clés seraient mis à jour. Il peut être intéressant de connaître la fréquentation de son site en installant des outils de statistiques automatiques sur les pages a priori les plus fréquentées (ou sur le site en entier si l'outil le permet) et également d'analyser les fichiers *log* disponibles sur le serveur et générés par le trafic. Cela remplit le double objectif de mesurer la rentabilité aussi bien que de recueillir des informations sur le comportement des visiteurs. On peut de ce fait améliorer son site et éventuellement négocier des espaces publicitaires.

On a donc vu que l'on peut améliorer l'indexation de ses pages et le référencement de son site grâce à la mise en place de consensus internationaux comme le Dublin Core. Mais, ceux-ci sont avant tout destinés à tenter de mettre en forme le Web et sont déconnectés des préoccupations financières qui envahissent peu à peu la politique d'Internet.

En effet, au vu de la conjoncture actuelle (augmentation du trafic, des délais de traitement, du volume des données, etc.) il est légitime de se demander si Internet ne va pas vers un référencement payant. Cette logique est déjà amorcée. En effet, la publicité est quasiment la seule source de revenus des outils de recherche. La multiplication de ces bandeaux publicitaires n'est pas sans effet : ralentissement du temps d'affichage, réduction du volume lisible sur l'écran, réticence des internautes, incompatibilité entre la publicité et le système d'exploitation de l'utilisateur. Malgré tout, il semble que l'on conserve une certaine intégrité dans le référencement.

D'autre part, les outils de recherche évoquent de manière de plus en plus systématique l'éventualité de faire payer leur service de référencement. Cela concernerait avant tout les sites à caractère commercial et cela jouerait sur la rapidité du traitement des soumissions et non pas sur le classement en lui-même. Les outils de recherche conservent néanmoins la volonté d'offrir un service gratuit pour la majorité des sites. Cette méthode pose des problèmes d'éthique puisqu'elle va à l'encontre des pratiques fondamentales de l'Internet, mais elle s'inscrit cependant dans une tendance actuelle qui fait de plus en plus payer les services. Se posent d'ailleurs les questions de crédibilité et de pertinence quant à la constitution des listes de classement.

Les logiques documentaire et financière qui régissent Internet évoluent pour l'instant en parallèle. Mais la situation est telle que ce ne sera pas toujours le cas et qu'il faudra trouver des compromis pour les faire coexister toutes les deux et les rendre interdépendantes.

II. Le site Lille3 Jeunesse

Dans le cadre du Master, j'ai effectué un stage pratique à la Bibliothèque des Bibliothèques du Service Commun de Documentation de Lille 3. Ce stage est de nature un peu particulière puisqu'il est la volonté d'un professeur d'université, M. Christian Looock, et part d'un projet préexistant basé sur les cours donnés dans le cadre du module de Littérature Jeunesse de l'U.F.R. I.D.I.S.T. Ce projet prend la forme d'un site Internet (Lille3 Jeunesse). La constitution d'une équipe de travail répondait cette année un besoin de pérennisation, de rénovation et d'enrichissement du site.

Cette équipe était constituée de sept membres dont moi, tous étudiants en Master S.I.D. première année. Une étudiante en D.E.A. a également proposé ses services. En outre, une documentaliste de Grenoble envoyait régulièrement des articles pour alimenter le site.

L'ampleur des tâches initialement prévues était trop importante. Certaines ont dû être mises de côté pour éventuellement être effectuées les années prochaines : Il semble difficile d'un prime abord de confier les acquisitions d'un département d'une bibliothèque universitaire à des étudiants de maîtrise, stagiaires de surcroît. Cette proposition n'a donc pu être satisfaite. Une opération de traitement documentaire a été amorcée, mais il reste un volume considérable de travail en attente. Là encore, il s'agit d'une mission fastidieuse et qui devrait normalement être assurée par des bibliothécaires en poste. Cela reflète une certaine politique entourant la littérature jeunesse au sein de l'Université : si le site a très bonne presse et que le fonds jeunesse est unanimement reconnu comme étant d'un grand intérêt public, l'Université ne se contente cependant que d'héberger le site et le fonds sans pour autant dégager de moyens humains. La survie du domaine de la littérature de jeunesse, qui est une des particularités de Lille 3, est donc remise en question chaque année et ne tient qu'à l'engagement et la motivation d'une petite poignée de personnes qui ne peuvent à elles seules assurer toutes les tâches. C'est pourquoi, le dépouillement régulier des revues professionnelles n'a pas pu être assuré. De même, il n'a pas été possible de couvrir tous les événements ayant trait à la littérature de jeunesse (par exemple le salon de Montreuil, ceux de Douai et de Valenciennes, etc.). Quelques unes des ambitions initiales ont cependant été réalisées : l'année a été riche en rencontres avec les auteurs, les illustrateurs et les médiateurs de la littérature jeunesse, des groupes de lecture en collège ont été mis en place afin de recueillir l'avis des enfants eux-mêmes sur les livres proposés.

Les activités des années précédentes ont bien entendu été maintenues : les étudiants de DEUST et de maîtrise (membres du groupe de stage inclus) ont continué à écrire des articles, des minithèses et des mags qui alimentent le contenu du site.

Le stage devait se dérouler en théorie à la BdB sur une demi-journée tout au long de l'année et sur une période bloquée en juin. Concrètement, le rythme a été totalement modifié. On s'est vite rendu compte qu'il fallait se montrer plus élastique. En effet, les tâches demandées ne pouvaient pas forcément être effectuées sur une demi-journée et à la BdB et la période bloquée de juin a essentiellement été consacrée au catalogage de certaines revues. Par contre, le suivi des missions toutes les deux semaines a été à peu près respecté. Il ne pouvait en être autrement pour conserver une unité au sein de l'équipe et pour réaffirmer régulièrement les orientations de travail et les délais à tenir.

Le stage devait aussi assurer la continuation du site pour un an. Tous les détails concernant cette tâche seront abordés plus loin dans le propos.

A. Présentation du cadre du stage

1. La littérature jeunesse à l'université

La littérature de jeunesse a longtemps fait figure de parent pauvre de la littérature. C'est pourquoi elle n'a intéressé que tardivement les universitaires. Quelques facultés l'ont incluse dans leurs programmes, mais chaque établissement ne propose qu'une vision parcellaire du domaine puisque la littérature est envisagée selon un thème précis. Par exemple, les cours dispensés par Jean Perrot à Paris XIII privilégient une approche ludiste en rapprochant la littérature de jeunesse du jeu. A Rennes, la littérature jeunesse est associée à l'image.

A Lille 3, la littérature de jeunesse est étudiée depuis 1992, mais elle connaît depuis des difficultés à subsister (cette formation de plus en plus réclamée se voit pourtant allouer de moins en moins d'heures). La formation proposée est pourtant tournée vers un objectif professionnalisant remis à l'ordre du jour par les réformes qui se succèdent. Les cours prodigués s'intéressent avant tout à la littérature contemporaine et à la production éditoriale actuelle.

Les travaux effectués à l'occasion de ces cours par les étudiants constituent la matière première qui fait vivre le site. Ils prennent la forme de soit de minithèses, soit de mags.

2. Le site Lille3 Jeunesse

a) L'historique

Le site a été créé en 1998 pour répondre à une volonté de diffusion des travaux des étudiants, dont le renouvellement est assuré chaque année par une nouvelle promotion. Il vient en remplacement à la méthode précédente qui s'est révélée fastidieuse. Le professeur chargé de l'enseignement avait en effet proposé de transmettre les travaux aux lecteurs de la revue *L'École Des Lettres* à laquelle il collaborait. La demande a été trop importante et il est vite devenu impossible d'y répondre de manière personnalisée. Il est donc apparu qu'il fallait un moyen de mise à disposition des travaux qui pourraient être utilisés par les personnes intéressées qui feraient elles-mêmes la démarche de collecter les informations.

Les rubriques ont évolué au cours des années, le seul noyau des minithèses demeurant inamovible. Par exemple, à partir de 1999, les « Sélections » avaient pour vocation d'assurer le dépouillement des revues professionnelles et les « Actualités » celle de tenir au courant des manifestations en rapport avec la littérature jeunesse. Toutes deux ont évolué dans le site actuel.

2000 est une étape importante dans la vie du site puisqu'elle voit la création de ce qui deviendra les « Mags » qui sont encore aujourd'hui une rubrique importante. On y présente les livres sous forme d'articles journalistiques, occasionnant de ce fait un élargissement du public, des professionnels de la documentation aux jeunes lecteurs.

Le site vit également sur les comptes-rendus de lecture menés depuis 1998 par un comité de lecture composé d'étudiants et d'enseignants du second degré.

Jusqu'en janvier 2000, ce sont des étudiants stagiaires qui ont fait vivre le site. A partir de cette date, une personne a été embauchée en Contrat Emploi Jeune (Marine Dormion) pour assurer la gestion du site et une partie de son alimentation.

Le poste laissé vacant par le départ de Marine Dormion à la rentrée 2003/2004 n'a pas été remplacé, suscitant de nouveau la crainte de voir le site périliter faute de personnes pour l'entretenir. Aucune intention de remplacer le poste n'ayant été manifestée, il a fallu faire appel à des étudiants, situation qui sera sans doute amenée à se répéter pour les années à venir.

Il y a deux ans, une stagiaire de D.E.S.S. a réalisé un audit du site. Ce dernier a révélé des chiffres très favorables quant à la fréquentation du site. En effet, celui-ci représente un quart du trafic total du site de l'université, ce qui est un score excellent lorsque l'on sait en plus qu'il n'existe aucun lien sur la page d'accueil et menant vers Lille3 Jeunesse. Le site reçoit le nombre important de mille visites par jour dont plus d'un tiers sont effectuées depuis l'étranger.

b) Le contenu

Le site est avant tout alimenté par les travaux étudiants répartis en plusieurs rubriques.

Les minithèses constituent le noyau central du contenu du site. Ce sont elles qui ont présidé à la décision de création du site. Elles se divisent en deux catégories : les auteurs et les thèmes. Ces travaux s'accompagnent d'une riche bibliographie qui en fait un outil appréciable pour les professionnels du livre et les personnes intéressées par le sujet.

Les mags sont des travaux de réécriture. Il s'agit de traiter de l'intrigue d'un livre sous une forme journalistique. Les étudiants s'organisent en groupes pour former une Une thématique qui présentera chaque travail comme autant d'articles. Les mags connaissent beaucoup de succès tant auprès des étudiants qui les écrivent qu'auprès des lecteurs, à tel point que le projet de délocaliser les mags sous une forme papier ou cd-rom a plusieurs fois été évoqué.

La rubrique Info Site est destinée à présenter le pôle de littérature de jeunesse de Lille 3 (la BdB, les acteurs, les orientations, ...). Elle est cette année complétée par les cours de M. Loock.

Les Bouts d'essai sont des travaux de DEUST basés sur le même principe que les minithèses réalisées, elles, par les étudiants de maîtrise.

La rubrique Journée Pro est un compte-rendu des journées professionnelles organisées régulièrement par de nombreux acteurs (bibliothécaires jeunesse du Pas de Calais, enseignants et formateurs IUFM, représentants du CRDP, de l'université, du Département de Formation aux Métiers du Livre et de la Documentation Médialille et de Jeunesse et Sports). Jusqu'à cette année la rubrique ne contenait que les actes officiels de ces journées. On y a inclus des articles d'étudiants sur la dernière journée mais aussi des articles sur d'autres manifestations comme les Parcours Professionnels Pour la Lecture Jeunesse.

La rubrique Sélections contient Biblio-Ado, une gigantesque bibliographie critique de romans pour adolescents, une nouvelle bibliographie thématique d'ouvrages théoriques (conte, illustration, auteurs classiques, bandes dessinées, médias), et enfin la Vie Des Livres qui consiste en la mise en ligne des critiques d'un comité de lecture mis en place par Marine Dormion et complétées par les réactions des enfants contactant le site par e-mail. Cette méthode de mise en relation des lecteurs avec le site a été remplacée cette année par un forum qui peine à trouver son public mais dont la nécessité est avérée.

Cette année a vu l'évolution de la rubrique des Liens ainsi que la naissance de la rubrique Rencontres qui devenait nécessaire face à la multiplication des comptes-rendus de rencontres avec les professionnels de la littérature jeunesse (auteurs, médiateurs, etc.).

B. Tâches de nature journalistique

Dans le cadre de mon stage, j'ai dû effectuer des comptes-rendus de manifestations professionnelles.

1. La 8^{ème} rencontre professionnelle à l'IUFM d'Arras

Le 27 novembre 2003 a eu lieu la journée professionnelle à l'I.U.F.M. d'Arras autour du thème « Télévision, vidéo, une alternative au livre ». Un suivi de cette manifestation semestrielle puis annuelle a été assuré au cours des années précédentes. Cette journée a rassemblé de nombreux professionnels dont Michèle Petit (anthropologue), Serge Tisseron (psychanalyste), Laurent Trémel (sociologue), Moka-Elvire Murail et Xavier-Laurent Petit (romanciers). La journée s'est achevée sur la présentation de livres sur le thème de la construction de soi.

Monsieur Loock, qui jouait le rôle de modérateur durant les débats, a fourni quelques jours auparavant les invitations et des places ont été réservées pour les stagiaires dans un bus affrété pour les étudiants de CAPES. Notre mission consistait essentiellement à assister aux conférences de Michèle Petit, Serge Tisseron, Laurent Trémel et à fournir un compte-rendu de leurs interventions qui serait par la suite mis en ligne. La charge de travail fut répartie entre les membres du groupe présents. Julien Siméoni et moi-même nous sommes chargés de l'exposé de monsieur Trémel, « l'univers des jeux vidéos ».

Nous avons rédigé un article conséquemment à cette journée. Il a été, une fois validé par M. Loock, mis en ligne sur le site Lille3 Jeunesse avec les autres productions des membres du stage.

Quelques semaines plus tard, M. Loock nous a fait parvenir un e-mail de réclamation signé de Laurent Trémel dans lequel il se disait insatisfait de ce qui avait été écrit. Suite à sa demande il a donc fallu corriger l'article²¹. Cela m'a fait prendre conscience que l'on n'est pas libre de mettre ce qu'on veut sur Internet et que tout ce qui y est publié peut rencontrer un vrai public. Rédiger un article de journal n'est pas anodin, il convient de se soucier de ce dont on parle, à qui on parle et des sources de l'information. Il faut également ménager les susceptibilités en cas de conflit.

2. Les Parcours Professionnels Pour la Lecture Jeunesse

Pendant deux jours, les mardi 10 et mercredi 11 février 2004, se sont tenus les Parcours Professionnels pour la Lecture Jeunesse au Grand Palais de Lille. Cet évènement de grande ampleur rassemblait tous les acteurs du livre jeunesse (auteurs, illustrateurs, éditeurs, médiateurs, ...). L'objectif fixé était de présenter tous les aspects de la littérature de jeunesse sous forme d'interventions, de la manière dont est fait un livre à la manière dont il est écrit et à sa promotion. Une quarantaine d'interventions de professionnels ont eu lieu suivant cinq formules : les tables rondes, les cartes blanches, les rencontres, les ateliers et les forums. L'équipe a eu la chance d'obtenir des passes, autrement coûteux pour le public, par l'intermédiaire de M. Loock. L'évènement était d'une telle ampleur que les membres présents

²¹ Voir Annexe 1 : l'article corrigé.

n'ont bien sûr pas pu tout couvrir. Il a fallu faire un choix préalable de sujets en adéquation avec le site et donc à traiter en priorité. Ce choix fait, les interventions ont été réparties.

Onze articles²² ont été mis en ligne. D'autre part, le Syndicat National de l'Édition (SNE), organisateur de ces journées, avait prévu de publier des actes des forums qui se déroulaient (au nombre de quatre). Nos services furent proposés pour qu'un maximum de forums puissent être retranscrits textuellement par les membres de l'équipe. Les retranscriptions²³ ont été effectuées après accord du SNE, elles pourront être mises en ligne sur le site Lille3 Jeunesse.

Ces journées furent l'occasion idéale de pouvoir côtoyer, pour la première fois pour la plupart d'entre nous, des acteurs renommés de la filière de la littérature jeunesse et de pouvoir enfin mettre un visage sur des noms cités maintes fois lors de cours théoriques.

C. « Nous Voulons Lire » : le jeu d'images

Pour le mémoire, M. Looock a souhaité que les étudiants produisent une minithèse sur un sujet de leur choix. Pour ma part, M. Looock m'a proposé un autre « contrat », d'autant que j'avais déjà rédigé une minithèse sur le thème des voyages vers les mondes parallèles durant l'année 2003.

M. Looock m'a présenté une série de quarante illustrations de quarante illustrateurs différents qui provenait du numéro anniversaire des 30 ans de la revue *Nous Voulons Lire !*. M. Looock a toujours été intéressé par la relation entre l'image et le texte ainsi que par le pouvoir dialectique de l'image. Il s'est dit que cela aurait pu être intéressant de mettre en scène de façon ludique ces illustrations afin que le joueur perçoive, en manipulant les images, que la manière de faire passer un message n'est pas le fruit du hasard. Une analyse trop théorique de l'image occulte selon lui la signification même de ces images.

Présentation de *Nous Voulons Lire !* :

La revue *Nous Voulons Lire !* a été créée en avril 1972 à l'occasion de l'Année Internationale du livre par Denise Escarpit, universitaire et chercheuse en littérature de jeunesse. Elle est reconnue par tous les professionnels du livre, éditeurs, auteurs, chercheurs.

Nous Voulons Lire ! est aussi une association créée en 1977 dont le nom est devenu *Nous Voulons Lire/CRALEJ* quand elle fut accueillie dans les locaux de la bibliothèque de Bordeaux. L'objectif était d'héberger la revue mais également des ouvrages (près de 20 000) afin de créer un centre de documentation.

La revue est depuis animée par sa fondatrice Denise Escarpit qui est épaulée d'un comité de bénévoles principalement composé d'enseignants. Cette revue est consacrée à la

²² Liste des articles que j'ai eu l'occasion de rédiger : *Texte et images*, Intervenants : Ronan Badel (auteur) et Hélène Kérillis (auteur) ; *Le choix éditorial dans la fiction*, Animateur : Mariette Darrigrand, Intervenants : Babeth Sebaoun (Bayard) - Caroline Westberg (Rageot) - Chloé Moncomble (Milan) ; *Comment sont faits les livres et à quel coût ?* Intervenants : Pierre Ducos (Nathan) - Anne-Marie Sarocchi (Gallimard) ; *La promotion pour faire connaître un auteur, un livre*, Intervenants : Xavier Decousus (Rageot) - Sandrine Dutordoir (Gallimard)

²³ *De la lecture partagée à la lecture autonome : les enjeux des premières lectures. Comment accompagner l'enfant dans sa démarche de lecture. Quels mécanismes, quelles réponses? Quelles méthodes sont identifiées et proposées dans le contexte de société actuel.* Échanges entre professionnels de l'enfance. Animateur : Dominique Guy Brassart (IUFM Lille). Intervenants : Juliette Campagne (assoc. Lis avec moi) - Marie-Agnès Gaudrat (Bayard) - Sophie Dubern (Conseillère) - Marie-Hélène Place (Hatier).

littérature jeunesse (comptes-rendus, entretiens, articles,...). L'image y a une place particulière. Elle y a toujours été utilisée comme un outil pédagogique, universitaire et familial. On peut déceler un intérêt particulier pour l'album.

Le jeu :

En me basant sur le cours de M. Loock intitulé « lire le texte, lire l'image, quelques points de méthodologie », je me suis efforcé de mettre en forme un petit jeu mettant en avant les différentes techniques de mises en scène de l'illustration utilisées pour véhiculer des informations. Chaque exercice met en avant un critère précis :

- « - le degré d'iconicité, c'est à dire la proximité par rapport au réel. De la photographie au dessin et au schéma, l'image chemine vers l'abstraction.
- les codes graphiques. Ils comprennent les lignes, les couleurs, la lumière, le respect ou non de la perspective. Si leur identification résulte d'une analyse parfois délicate, leur effet sur le lecteur peut être instantané et inconscient
- les techniques mises en œuvre. Le recours à des supports, à des pigments, à des outils d'application différents peut produire des effets différents.
- les valeurs de cadre. Elles englobent les plans (du panoramique au gros plan et à l'insert), les angles de prise de vue (normal, plongée, contre plongée, champ-contre-champ), les points de vues (frontal, arrière, de côté).
- l'insertion dans la page et le document
- la relation texte-image. »²⁴

J'ai tout d'abord tenté de mettre en œuvre un exercice ayant pour consigne « compléter la suite d'images proposées en en choisissant une parmi les autres proposées ». La mise en œuvre de ceci a été faite sur une base HTML comme si l'on créait un site Internet. Ce format est par la suite très pratique si l'on veut modifier le support des fichiers (sur cédérom ou en ligne).

J'ai également réfléchi à d'autres types de jeux possibles. En voici certains : un puzzle où il faut remettre les pièces à leur place à l'aide de la souris, une image « floutée » que l'on doit reconnaître en la retrouvant parmi plusieurs autres images proposées, un quiz avec des questions précises sur chacune des images, un *memory* (trouver une paire identique d'images parmi plusieurs images retournées et disposées aléatoirement), une image cachée qui se révélerait petit à petit de façon morcelée et qu'il faudrait reconnaître le plus rapidement possible parmi un choix d'images...

Techniquement, je me suis vite rendu compte que face à mes capacités de programmation, j'allais devoir me limiter au jeu le plus simple (compléter une suite) car les compétences n'arrivaient pas à suivre les idées. Je me suis alors lancé dans la création de pages HTML. Cependant, la structure du jeu s'est vite révélée problématique pour une programmation HTML. En effet, un choix de plusieurs images était proposé pour compléter chaque image. Chaque possibilité générerait une page, multipliant leur nombre de façon telle que j'ai dû trouver une autre solution.

C'est alors que M. Loock m'a fait découvrir Hot Potatoes 6, un logiciel spécialisé dans la création de jeux interactifs en ligne basés sur le langage HTML et JavaScript, mais dont l'utilisation ne nécessite pas de connaissances particulières. Cinq types de jeux sont proposés par ce logiciel : un quiz à choix multiple, des phrases à trous à compléter, des mots croisés,

²⁴ Voir le cours de M. Christian LOOCK, « lire le texte, lire l'image, quelques points de méthodologie ».

des mots mélangés à remettre dans l'ordre. Cela convenait parfaitement à ma tâche et l'utilisation vite assimilable m'a convaincu d'abandonner la conception en HTML pour l'utilisation de Hot Potatoes 6.

D. Tâches de nature documentaire

La période bloquée de juin a été consacrée au catalogage de revues restées en souffrance à la BdB. Il s'agissait entre autres de numéros des périodiques *Citrouille*, *Griffon*, *Lire au lycée*, ... J'ai eu pour mission de rentrer les derniers numéros de *Lire au collège* (une petite dizaine de numéros environ). La tâche à accomplir était le dépouillement d'articles et leur indexation dans la base littérature de jeunesse qui gère le fonds de la BdB et qui utilise le logiciel BCDI. Il a fallu attendre juin pour que chaque membre de l'équipe ait des disponibilités accrues. Un problème d'informatique est vite apparu puisque le système informatique de l'Université ne supportait que l'occupation de deux postes uniquement pour l'application BCDI. Cette difficulté ne fut résolue qu'en juin, mais c'est alors un manque d'ordinateurs qui a nécessité la mise au point d'un planning et une petite réorganisation matérielle du bureau de la BdB.

Cette tâche ne s'est pas révélée des plus aisées car les moyens débloqués en termes d'heures de cours par l'UFR ne permettent pas une formation poussée à BCDI, dont l'utilisation ne s'avère d'ailleurs pas d'une simplicité extrême. Le manque de temps n'a permis que l'alimentation de la base et pas une révision du fonds déjà enregistré.

E. L'entretien du site

Le principal travail qui nous fut confié à Julien Siméoni et à moi par M. Loock est l'entretien technique du site. La question de donner un an de sursis au site s'est posée après le départ de Marine Dormion qui en assurait seule la gestion. La volonté de M. Loock était de recruter des stagiaires capables d'assumer l'entretien du site parmi les étudiants de Maîtrise SID. Les principales tâches étaient d'alimenter Lille3 Jeunesse de nouveaux textes produits au cours de l'année, mais aussi de procéder à une révision technique du fonctionnement et une révision graphique du site. Le travail s'est fait avec les logiciels Dreamweaver et FrontPage.

Une partie du stage devant s'effectuer *in situ*, un accord a été trouvé avec Médiatic pour une mise à disposition d'un moyen informatique dans leur salle. Cela s'est montré utile pour installer régulièrement une mise à jour du site à partir de leur poste informatique. De plus, M. Loock a ainsi pu intervenir directement dans le site. Durant la première moitié du stage, il fallait travailler séparément sur la partie qui nous incombait chacun, puis faire une mise en commun des modifications pour enfin porter une version du site remise à jour au CRI (Centre de Ressources Informatiques) chargé de mettre en ligne la dernière version du site. Mais, pour la deuxième moitié du stage, le CRI ayant manifesté une certaine lassitude à remettre régulièrement le site à jour, les codes d'accès au serveur nous ont été confiés pour qu'il soit possible de s'acquitter de cette tâche depuis nos ordinateurs personnels.

1. Mes tâches

J'ai eu à m'acquitter de différents travaux pour le site. En voici le panel en détail.

a) La bibliographie thématique

Marine Dormion m'a tout d'abord confié la mise en ligne d'une nouvelle série de bibliographies qu'elle avait préparée. C'est une bibliographie thématique (sur les généralités, le conte, l'illustration, les auteurs classiques, la bande dessinée, les médias) d'ouvrages théoriques sur la littérature de jeunesse. Cette sélection a été effectuée à partir du fonds de la Bibliothèque des Bibliothèques, les notices concernées ont donc été extraites de la base informatique de la BdB. La difficulté de la tâche, en plus de la création de cette nouvelle rubrique, résidait dans le fait que les notices extraites n'avaient pas été remises en page et qu'il fallait donc vérifier qu'il n'y ait pas de fautes ou de caractères parasites.

b) Le Mag Garçons / Filles

Cette année, les mags ont été réalisés d'après une commande passée par le CRDP de Grenoble, à savoir traiter des relations entre les garçons et les filles. Les étudiants du module de littérature de jeunesse ont donc répondu à la demande et une dizaine d'articles a rapidement pu être mise en ligne, d'autres devraient suivre. Cette commande correspond avec la sortie d'un recueil de nouvelles, *Des filles et des garçons* chez Thierry Magnier, dont les étudiants se sont fortement inspirés pour rédiger leur mag.

Il fallait mettre en forme ce mag en accord avec la charte graphique qui régit les mags préexistants sur le site. Il s'agissait donc de conserver la forme d'un magazine qui avait été attribuée mais aussi d'utiliser des couleurs rappelant celles de l'université (bleu clair, jaune clair et blanc). La page doit être dynamique et proposer plusieurs entrées : par thème (ou rubrique), par titre de roman ou même directement par le châpo de l'article.

D'un point de vue strictement personnel, j'ai rédigé au cours de l'année 2003 un mag qui s'inscrit dans la rubrique des mondes parallèles.

c) La Vie des Livres

La Vie des Livres est la rubrique dans laquelle sont publiés les comptes-rendus du comité de lecture créé et animé par Marine Dormion. Il s'agit de critiques de livres récemment parus faites par des étudiants, des enseignants du second degré et des professionnels du livre. Les nouvelles critiques ont afflué dès le début du stage puisque le comité de lecture a repris le 5 novembre 2003.

Il me fallait gérer cette partie du site à l'aide d'une fiche technique envoyée par Marine dans laquelle elle explique les règles à appliquer pour toute nouvelle critique à mettre en ligne (elle insiste sur l'importance des métadonnées par exemple). De plus, elle proposait des modifications à apporter en vue d'une amélioration, essentiellement une révision de l'archivage des anciennes critiques. Il s'agissait essentiellement d'une collaboration avec Marine Dormion puisque c'est elle seule qui prenait cette partie en charge aussi bien dans son alimentation que dans sa gestion.

J'ai d'abord corrigé tous les liens cassés qui préexistaient à mon arrivée, ancrs nommés incluses, une par une avec Dreamweaver.

Un tri dans les archives a été nécessaire : j'ai réorganisé les pages d'index, c'est-à-dire que j'ai fait passer les critiques apparaissant dans l'index titres vers l'index archives dans lequel elles devaient normalement se trouver. Cette translation a également nécessité le transfert de tous les fichiers HTML concernés du répertoire « articles » vers le répertoire « archives » (situés tous deux dans le répertoire « selections/comite/ »). Cela a engendré une correction très importante au niveau des liens cassés par le changement d'adresse des pages induit par le changement de répertoire. Cependant cette réorganisation coûteuse en temps passé à rétablir les liens s'avérait indispensable car il n'y avait plus aucune logique repérable dans la répartition entre les archives et les articles. Pour corriger les liens, j'ai utilisé le détecteur de liens cassés du logiciel Dreamweaver et j'ai modifié « à la main » l'adresse à l'endroit indiqué, ce qui s'est avéré fastidieux (plusieurs centaines de liens à réparer).

Une fois le tri effectué, j'ai réorganisé la page d'archives en rangeant tout par ordre alphabétique et en instaurant un système d'ancres nommées permettant de naviguer directement vers l'une des catégories de livre (album, documentaire, bandes dessinées, fictions) car la page devenait trop longue pour être lisible. Le répertoire « archives » comporte actuellement au moins 300 fichiers qui sont autant de critiques de livres.

J'ai mis en ligne les nouvelles critiques (au moins 76). Pour chacune, il a fallu créer une page à partir du modèle fourni par Marine Dormion, ainsi qu'aller chercher sur un site de vente en ligne de livres (comme Alapage ou Amazon) la photo de la première de couverture, la renommer correctement, la placer dans le répertoire « images », la redimensionner et l'insérer dans la critique. Chaque nouvelle critique de livre doit comporter un lien vers les index titres, auteurs, accueil et éventuellement théma si ce livre correspond à un des théma existant (fantasy, vacances, expression de soi, ah... l'amour, suspense et cie) et se voit adjoindre un petit N rouge pour signaler la nouveauté. Inversement, il faut assurer la réciprocité des liens et aller dans chaque index pour faire pointer la nouvelle entrée vers la critique. A chaque nouvelle création il faut garder en tête de donner aux fichiers des noms compréhensibles et penser à renseigner un minimum les balises méta ou tout au moins donner un titre clair à la page (titre du livre).

Au nombre de thémas déjà existants s'est ajoutée une nouvelle rubrique thématique : le théma « historique ». Il a de nouveau fallu veiller à instaurer de nombreux liens depuis les livres concernés vers la page théma et inversement. Chaque théma a sa propre couleur qu'il donnera à la critique des livres qu'il comprend. Pour la création de ce nouveau théma, j'ai dû modifier les couleurs des critiques des livres qui se classaient maintenant dans le théma « historique » et ajouter sur la page critique une mention claire du nom du théma auquel elle est liée.

Enfin, il a fallu refondre la page de sommaire à l'aide des indications fournies par Marine Dormion en y insérant les liens nécessaires vers le nouvel éditio (les éditos précédents ayant été transférés également dans les archives) et vers les nouvelles critiques.

A l'origine, il fallait également gérer l'ajout des avis des enfants parvenus par e-mail aux critiques des livres correspondants ainsi que traiter toutes les questions et demandes reçues de la part des visiteurs. Malheureusement, les messages qui étaient en attente sur le serveur n'ont pu être récupérés et toutes les informations qui y étaient contenues ont été définitivement perdues. Seuls quelques avis d'enfants préalablement mis de côté ont pu être utilisés et mis en ligne. Cet incident a mis en lumière la nécessité d'installer un forum qui mettrait en relation les visiteurs entre eux mais aussi avec les gestionnaires du site. La décision prise, il a fallu rapidement couper tous les liens donnant l'adresse e-mail du site pour éviter que les gens l'utilisent inutilement. Dans le cas de La Vie des Livres, cette mission a encore une fois été fastidieuse : sur chaque page de critique se trouvait un lien vers l'adresse e-mail complétée du nom du livre (par exemple : <mailto:bdblj@univ-lille3.fr?subject=infante>). On ne pouvait donc pas demander à Dreamweaver d'effacer tous les

liens d'un coup car chacun avait sa propre spécification. En conséquence, j'ai donc dû effacer presque tous les liens un à un.

d) La navigabilité

Le travail effectué pour la Vie des Livres au niveau des liens et de l'arborescence des fichiers a été étendu à l'ensemble du site. En effet, on a constaté que le site prenant de l'ampleur, la gestion des fichiers est vite devenue anarchique. Il a été nécessaire d'y remédier. J'ai par conséquent essayé d'évaluer et de repérer les liens cassés afin de les réparer. Ce fut assez payant puisque une bonne centaine de liens était à refaire. Personnellement, je me suis attaché à nommer et à renommer correctement les fichiers. En effet, les noms de fichier ne doivent comporter aucune majuscule, aucun accent, aucun espace... et être le plus court et le plus intelligible possible. Dans le cas contraire, des problèmes de navigation (comme des erreurs 404) peuvent survenir.

J'ai ensuite fait un travail de détection et d'effacement des fichiers orphelins du site. Un fichier orphelin est un fichier séparé des autres car il ne mène à rien et rien ne mène à lui. Configuré de la sorte, il n'est d'aucune utilité. Avec l'outil de détection de Dreamweaver, j'ai recherché ces fichiers puis je les ai étudiés un par un pour déterminer lesquels devaient être réintégrés et lesquels devaient être définitivement effacés. 169 fichiers ont pu être effacés, ce qui correspondait à environ 7% du volume total du site, soit 3 Mo qui occupait inutilement le serveur.

Tout nouvel ajout ne se fait pas sans contrainte et il convient de respecter la charte qui régit le site. Les normes minimales nous ont été fournies par Marine Dormion. Elles concernent la police qui devait être de type Arial, de taille 2. Les couleurs ne devaient pas être attribuées aléatoirement : chaque partie du texte possède son code de couleur (texte = #000078, lien = #0000FF, lien actif = #800080, lien activé = #FF00FF).

2. Tâches partagées avec Julien Siméoni

Des tâches plus quotidiennes étaient aussi à assurer. Des mises à jour régulières étaient souhaitables pour que le site soit vu comme dynamique et prenne une forme un peu plus marquée de revue. Ces tâches de mises à jour régulières étaient partagées entre Julien Siméoni et moi-même. Elles consistaient en la mise en ligne de l'édito, de nouvelles minithèses, de bouts d'essai, de comptes-rendus des journées professionnelles, des rencontres entre les professionnels du livre et les membres de l'équipe, des cours de M. Loock sur la littérature jeunesse, etc.

La mise en ligne d'un nouvel article signifiait à chaque fois de respecter la charte de mise en page ainsi que l'intégration de ce document au site par la mise en place de liens pointant vers d'autres pages et réciproquement (généralement, on installe au moins un lien vers l'accueil et vers la page de sommaire de la rubrique concernée).

Cette année a vu la naissance de nouvelles rubriques. Nous avons mis en place, entre autres, une rubrique Rencontres. Il a par conséquent fallu déplacer certains fichiers d'articles existants vers le répertoire spécialement créé pour la rubrique. Ce qui entraîne, comme déjà vu, les problèmes de liens cassés à rectifier. D'autres rubriques sont apparues comme, dans Info Site, les archives d'édito, les cours de M. Loock, et dans Journées Pro, les Parcours

Professionnels pour la Lecture Jeunesse. Tout ceci s'est fait en plus de l'enrichissement de toutes les autres rubriques existantes.

Julien Siméoni a mené, quant à lui, en plus de ce que je viens de citer, un travail d'ergonomie sur le site. Il y a eu une refonte esthétique du site avec la retouche et la mise en place sur une nouvelle page d'accueil dynamique des dessins de Coralie Jéret qui confèrent au site une présentation plus moderne et attractive. Julien a en outre veillé à l'installation du forum et d'un moteur de recherche.

Travailler à plusieurs sur un site ne va pas sans mal et demande une coordination efficace et effective. En effet, il est régulièrement arrivé, du fait d'être « bicéphales » sur une même tâche, même différenciée, qu'il y ait des problèmes de synchronisation. Il se peut aussi que le même travail soit effectué en double à cause d'initiatives malheureuses et d'oublis lors des communications, c'est alors le premier qui l'a achevé qui le mettra en ligne. Il peut en résulter parfois pour le *surfer* une impression de manque de cohésion, rarissime il est vrai car nous sommes restés vigilants, mais néanmoins inévitable.

F. Application de la partie théorique au site Lille3 Jeunesse

L'application au site Lille3 Jeunesse de tout ce qu'on a pu voir et expliquer au cours de la première partie de ce mémoire fournit assez de travail pour constituer un stage à part entière. Comme nous l'avons déjà écrit, un temps raisonnable pour mettre en place une indexation des pages d'un site et leur référencement sur les outils de recherche représente au moins trois mois. Je me propose ici de faire un premier dégrossissement de ce qui peut être appliqué pour ce site.

Il faut tout d'abord prendre en considération certains éléments. A l'origine de la création du site, il n'y avait aucune véritable perspective de communication, en dehors de fournir aux gens qui le souhaitaient des minithèses. Il n'y a donc pas eu de véritable cahier des charges strict sur lequel les futurs gestionnaires du site auraient pu se reposer, notamment pour l'indexation des pages. En ce qui concerne la facilité d'accès au site par le public, il est tout à fait injuste de ne pas pouvoir placer un lien sur l'accueil du site de l'université (ou tout au moins à l'intérieur d'une des rubriques) alors même que Lille3 Jeunesse y est hébergé ! Sans oublier qu'il est d'une qualité honorable et qu'il fait un quart du trafic total du site de l'université. Cela pourrait constituer une première reconnaissance du travail effectué et par les professionnels de Lille 3 et par les étudiants de l'U.F.R. I.D.I.S.T.

1. Le Dublin Core

Nous avons vu que dans un souci de respect des normes documentaires, il serait bienvenu d'intégrer à Lille3 Jeunesse les balises spécifiques du Dublin Core. Comme constaté dans la partie I, le Dublin Core n'est qu'un support et il évolue en fonction du contexte d'application. Il peut donc être modifié selon le « métier » pour lequel il est employé.

a) L'Éducation Nationale et le Dublin Core

Les initiatives définissant les métadonnées sont rarement appliquées au domaine de l'éducation, tout du moins pour les sites francophones. Une tentative d'adaptation du Dublin Core a été produite par un petit groupe de travail du CRDP de Montpellier qui s'est réuni huit fois au cours de la période décembre 2000 – avril 2001. Ce groupe a été constitué dans le cadre d'une mission pour les ressources en ligne de l'académie de Montpellier, suite au constat de l'absence de schéma normalisé de métadonnées adapté au système éducatif français. Ce groupe a finalement sélectionné douze éléments parmi ceux proposés par le Dublin Core et les a adaptés aux ressources éducatives françaises. Il en a créé sept autres pour préciser la portée éducative des ressources. Ces sept éléments sont pour l'instant nommés GTME (Groupe de Travail Métadonnées Éducation).

Voici les éléments retenus : pour le Dublin Core : titre, auteur, sujet et mots-clés, description, éditeur, date, format, type, identifiant, relation, droits, langue, pour le GTME : statut, niveau, discipline, nature, interactivité, public.

Le statut est « l'état de la ressource dans son degré actuel d'achèvement », le niveau est la « classe ou niveau d'enseignement des utilisateurs finaux de la ressource », la discipline est la « discipline(s) ou domaine(s) pouvant être concerné(s) par la ressource », la nature est la « forme de la ressource, déterminée en référence au contenu intellectuel de celle-ci », l'interactivité est le « type d'interactivité induite par l'utilisation de la ressource » et enfin le public est la « catégorie d'utilisateurs à laquelle la ressource est principalement destinée ».²⁵

b) La syntaxe des métadonnées

Dans le code, les métadonnées du Dublin Core seront insérées comme précisé dans le Dublin Core. Les métadonnées GTME auront une syntaxe approchante, par exemple :

<META NAME= « GTME.Status » CONTENT= « document de travail »>

Après avoir normalisé les métadonnées, il faut aussi normaliser leur contenu. Pour ce faire on utilise un schéma d'encodage, soit les listes d'autorités établies par le CNDP, soit la proposition du GTME dans le cas qui nous intéresse.

Le GTME et le Dublin Core sont des outils qui me semblaient adaptés pour procéder à l'indexation des pages du site Lille3 Jeunesse. Ce site propose en effet de nombreux textes pédagogiques et éducatifs qui entrent tout à fait dans le cadre du GTME. Je n'aurai malheureusement pas le temps de me pencher d'avantage sur ces listes pour choisir les mots-clés les plus adéquats pour le site. Cette opération doit par ailleurs faire partie intégrante de l'indexation qui est une tâche à accomplir de manière suivie.

²⁵ Voir l'introduction au GTME sur le site de l'Académie de Montpellier : <http://www.ac-montpellier.fr/ressources/IntroGTME1-0.doc> (ainsi que GTME1-0.doc et GTME1-1.doc).

2. Les mots-clés

a) Le contexte

L'audit réalisé il y a deux ans, a permis de dégager le public fréquentant le site. On a constaté que la majorité des utilisateurs appartiennent à l'Éducation Nationale. On remarque aussi qu'un tiers des visites viennent d'étudiants étrangers. Les enfants sont également de réguliers visiteurs. Le site touche de nombreux secteurs du livre et de la lecture (l'enseignement initial et continu, la création, la critique, la lecture publique, la lecture privée et l'édition). Dans ces conditions, la cible est avant tout française. Le problème des étudiants étrangers pourrait nous amener à réfléchir sur l'éventualité de la mise à disposition d'une traduction au moins partielle du site (au minimum la page d'accueil). En fonction des ambitions futures du site, on pourrait donc mettre en ligne une, voire plusieurs, « page(s) alias ». Mais quelle langue choisir ? Si le public étranger se constitue d'étudiants, pourquoi ne pas utiliser l'anglais, langue universitaire par excellence ?

Les pages que l'on pourrait choisir de référencer seraient la page d'accueil, puis les pages d'introduction aux rubriques « Infos Site », « Les Mags », « Minithèses », « Bouts d'essai », « Journées pro », « Rencontres », « Liens ». En ce qui concerne la rubrique « Sélections », son cas est un peu particulier car elle contient plusieurs sous-rubriques dont la page de sommaire doit être proposée à l'indexation : « la Vie des Livres » et « Biblio Ado ».

b) Composer la liste

En fonction de ce que l'on vient de voir, il serait judicieux de réfléchir à quelques mots-clés en anglais, en plus de ceux que l'on va dégager. Ceux-ci pourraient être inclus dans les « pages alias » si elles sont créées, mais pas forcément. Il peut en effet être utile d'inclure des termes anglais pour dans le code des pages françaises pour qu'une requête tapée en anglais donne Lille3 Jeunesse en réponse, même dans le cas où le site est unilingue.

En suivant les prescriptions énoncées auparavant pour la constitution de la liste de mots-clés, il faut établir une liste à partir du texte de la page que l'on souhaite indexer en prenant les termes qui reviennent le plus fréquemment. Le problème des pages proposées précédemment est qu'elles sont assez pauvres en texte. Ce n'est donc pas sur ces pages-mêmes qu'il convient de chercher les mots-clés adéquats, mais sur la rubrique dans son entier, ce qui représente une somme de travail conséquente et ingrate car peu de termes seront finalement retenus. A cette liste statistique s'ajoute une liste éditeur. Pour cette dernière, il faudra demander à M. Loock de fournir dix mots-clés qui décrivent selon lui le mieux la partie du site indexée. La liste des concurrents se compose des mots-clés les plus inclus dans leurs métabalises. Une liste intuitive viendra en complément. Aucune étude et aucune statistique n'entre en ligne de compte, il s'agit d'utiliser les termes dont on pense qu'ils seront le plus employés par le public. Dans le cas de Lille3 Jeunesse, un problème se pose : le panel des visiteurs est en effet étendu et il semble qu'il faut faire coexister des mots-clés de différents niveaux pour répondre à tous les publics.

c) Classer les mots-clés

Le principal problème de cette tâche réside à nouveau dans la pluralité des publics du site. Comment dès lors classer les mots-clés sans pénaliser un public ? Faut-il orienter l'indexation de certaines parties vers un public et d'autres parties vers un autre public ? Mais cela reviendrait à morceler le site et à revoir la conception de Lille3 Jeunesse, ce qui n'est sans doute pas une des priorités.

d) L'insertion des mots-clés dans la page

Les mots-clés précédemment dégagés répartis de façon cohérente dans la balise <TITLE> et les métabalises *Keywords* et *Description*, il faut penser à la mise en place de ces mêmes mots ou d'autres mots-clés dans la page pour optimiser l'indexation par les moteurs de recherche. Il faut bien veiller à utiliser des mots-clés comme titres dans les textes. Comme il y a peu de texte dans les pages d'accueil mais qu'elles comportent toujours un dessin de Coralie Jéret, il est important également de lui attribuer des mots-clés. Ceux-ci doivent présenter de manière claire le contenu des parties que le dessin illustre.

e) Piloter les robots

Pour cette étape, il faut se demander quelles sont les pages du site qu'il faut censurer à l'indexation par les moteurs. A priori, je ne vois aucune page susceptible d'être cachée des robots. Le site a pour objectif de faire connaître tous ses travaux. Le seul cas qui se présenterait serait de vouloir faire obligatoirement passer un visiteur par la page d'accueil du site ou d'une rubrique et donc de lui interdire l'accès direct à une page en ne la soumettant pas à l'indexation. Autrement, aucun impératif pécuniaire ou commercial n'exige de garder une certaine confidentialité.

3. Le référencement du site

Le référencement des pages doit être fait en complément de leur indexation. Il nécessite une grande préparation C'est une étape qui intervient une fois les mots-clés établis. Elle doit être réalisée de manière cohérente et rigoureuse. Je n'aurai malheureusement donc pas la possibilité de la prendre en charge.

a) Réviser le code HTML des *frames*

Le problème majeur du site est l'utilisation de *frames* ou cadres. En effet, même si de plus en plus de moteurs arrivent à les prendre en compte, les résultats que l'on obtient lorsqu'on fait des recherches avec eux sont hasardeux. Il est important de réfléchir soit à une nouvelle méthode de présentation du site, soit, plus simplement, à retirer les *frames* des pages d'accueil et de présentation et d'y inclure à la place des liens de navigation disposés dans la

page. Si les *frames* sont conservés, il faut essayer d'optimiser le code HTML des pages qui les mettent en forme. Il faut alors se baser sur l'exemple donné en annexe 9 pour la réécriture du code.

b) Le choix des outils de recherche adéquats

Le public qui visite le site est essentiellement français, il faut orienter les référencement vers des outils de recherche francophones avant tout. On ne connaît pas la nationalité des étudiants étrangers et d'ailleurs, il n'y a aucun pays qui ressorte de manière marquée. Il s'avère donc inutile de solliciter un outil de recherche d'une nationalité donnée. Si on décide de prendre l'anglais dans le cas où des pages du site seraient traduites, on se « contenterait » des moteurs de recherche internationaux (qui seraient utilisés de toute façon même dans le cas d'un site exclusivement unilingue).

Le site Lille3 Jeunesse a en outre une forte identité régionale, on peut donc l'inscrire dans un moteur de recherche du Nord s'il en existe.

Il pourrait être intéressant de rechercher un annuaire spécifique idoine pour s'y inscrire. La recherche peut être effectuée grâce aux adresses données en bibliographie.

Lille3 Jeunesse est un site à visée documentaire qui se veut extrêmement rigoureux dans sa publication. Les mises en ligne ont reçu en amont l'accord d'universitaires. Les sujets traités le sont avec sérieux et de manière complète. De ce constat, il ressort que l'on peut tenter de se faire connaître des « Signets de la BNF ». Une inscription ne serait pas seulement prestigieuse, elle apporterait une crédibilité supplémentaire au site, avec une assurance pour les visiteurs de trouver une information complète et sérieuse.

c) La phase de référencement

Pour le référencement automatique, utiliser le logiciel ProMoteur est ce qui semble le plus souhaitable étant donné qu'il est francophone et qu'il propose donc le référencement aux moteurs francophones. Si l'on souhaite pousser plus loin le référencement à un niveau international, le logiciel SubmitWolf peut remplir cet office.

Cela ne doit pas faire oublier le référencement manuel, plus efficace. Pour accomplir cette mission au mieux, il ne faut pas oublier de mettre toutes les informations utiles de côté.

4. Les autres techniques de promotion

Comme vu précédemment, on peut assurer la promotion de son site à l'aide de l'échange de liens pour lequel il est nécessaire d'établir une liste des sites traitant du même sujet que le Lille3 Jeunesse. On peut les trouver en formulant des requêtes avec les mots-clés choisis pour le site dans les différents outils de recherche. Il est aussi possible de détecter les liens qui pointent vers le site en utilisant la commande « link : <http://www.univ-lille3.fr/jeunet> » dans les principaux moteurs de recherche. En visualisant la liste de résultats, on peut procéder à un tri pour mettre en place un échange d'adresses officiel.

Le *Webring* peut être un système intéressant, mais il s'utilise surtout pour des sites personnels à caractère non officiel (comme des sites de fans).

La *newsletter* présente par contre un intérêt qui n'est pas à négliger. Il serait intéressant de repérer une newsletter qui correspondrait au thème du site et de s'y faire connaître, mais il serait bien plus gratifiant et fidélisant de créer une *newsletter* pour le site, surtout si l'on souhaite un développement vers un système de mises à jour régulières. C'est une tâche ardue qui demande une gestion et un suivi sérieux, mais elle fidélisera de manière plus sûre les visiteurs du site.

La conférence de presse virtuelle est envisageable, mais elle n'est pas indispensable.

La première partie du mémoire citait en outre les campagnes de publicité virtuelles. Il ne me semble pas que ce soit un procédé applicable avec Lille3 Jeunesse. Le site n'a en effet aucun but lucratif, il n'a pas les ressources financières nécessaires pour supporter un tel coût, et, plus important, cela nuirait à l'intégrité intellectuelle qui a été le mot d'ordre jusqu'à présent.

Bien gérer une politique de promotion demande de connaître la fréquentation de son public au moyen d'outils statistiques. Il n'a malheureusement pas été possible d'obtenir d'aide du CRI qui n'a pas été en mesure de fournir les fichiers *log* nécessaires à une telle démarche. Il est souhaitable, vu la situation, d'envisager l'installation d'outils statistiques sur le site. Ceux-ci permettraient de connaître un peu mieux la population et la fréquence des visites.

Il faut se référer à la bibliographie pour trouver les adresses des sites relatifs aux différentes méthodes évoquées.

Les missions qui m'ont été confiées cette année étaient suffisantes pour un seul stage. M. Loock a émis le souhait de voir le site trouver son public. Il est vrai qu'il mérite que l'on mène une politique de référencement que j'aurais aimé effectuer. Cependant, la tâche est d'une ampleur telle qu'elle demande qu'un autre stage y soit consacré. Je pense néanmoins avoir apporté quelques bases qui pourront aider la personne qui s'occupera éventuellement de ce travail.

Bien que je regrette de ne pas avoir pu référencer le site par moi-même, j'ai fortement apprécié la diversité des tâches qui m'ont été confiées même si elles étaient beaucoup tournées vers le site, ce que j'ai voulu. Ce fut une occasion unique de s'occuper d'un site traitant de la littérature de jeunesse avec autant de liberté et c'est sans doute la seule opportunité de traiter d'un sujet aussi intéressant dans le cadre universitaire. J'ai également été sensible à la prise en considération des travaux effectués par les étudiants.

Conclusion

Être webmaster, ce n'est pas seulement mettre du contenu en ligne, c'est aussi assurer à son site une promotion et une compétitivité qui justifient sa création. Cela passe par l'indexation et le référencement du site pour lequel il travaille, ces deux opérations étant soumises à des marches à suivre qui les rendent plus efficaces. Promouvoir son site dépend avant tout de soi et de la rigueur que l'on y met mais il serait utopique de croire que l'entraide rencontrée sur Internet n'est d'aucun secours. Les sites partenaires ou concurrents sont d'une aide précieuse lorsqu'il s'agit d'entreprendre une politique de communication vers l'extérieur.

Il est vrai que, malgré une portée documentaire qu'ils doivent conserver, l'indexation et le référencement sont devenus des opérations de marketing qui se préparent longuement et sérieusement et qui nécessitent une gestion et un suivi rigoureux. Il faut utiliser tous les moyens autorisés mis à disposition pour augmenter le classement de son site dans les outils de recherche dont dépend sa survie.

Mais être webmaster, c'est aussi assurer des missions à caractère plus journalier. Il faut posséder des compétences techniques pour assurer son bon fonctionnement. Il est de la responsabilité d'un webmaster d'instaurer une ergonomie agréable et une navigabilité aisée, deux critères nécessaires pour espérer une bonne fréquentation et une fidélisation des internautes. En tant que gestionnaire, il ne peut être que souhaitable qu'il ait une bonne vision d'ensemble parallèlement à une vision précise de chaque partie de son site. Il doit donc en connaître l'arborescence, les rubriques, ... Afin qu'il conserve une idée exacte de l'orientation à donner au site, il peut s'avérer judicieux que le webmaster soit aussi un producteur de contenu quand cela lui est possible ou tout au moins qu'il s'intéresse à ce qu'il met en ligne.

Ces préceptes sont ceux que je me suis efforcé de mettre en œuvre avec le site Lille3 Jeunesse dont j'ai assuré une partie de la gestion. Aux tâches purement techniques se sont ajoutées des tâches plus « intellectuelles » qui ont permis d'assurer plusieurs stades de la mise à disposition de l'information.

Bibliographie

- ESKENAZI Jean-Pierre F., *Référencement : comment référencer son site Web*, Versailles : Webedition, 1999.
- FOENIX-RIOU Béatrice, *Recherche et veille sur le Web visible et invisible : agents intelligents, annuaires sélectifs, interfaces des grands serveurs, portails thématiques*, Paris : Bases, Paris : Éditions technique et documentation, 2001.
- GUICHARD Éric (sous la direction de), *Comprendre les usages de l'Internet*, Paris : Éditions Rue d'Ulm, 2001.
- KING Andrew B., *Optimisation de sites Web*, Paris : CampusPress, 2003.
- LEFEVRE Philippe, *La recherche d'informations : du texte intégral au thésaurus*, Paris : Hermès Science Publications, 2000.
- PUBLICITYWEB, *Référencer et promouvoir son site*, Paris : Micro Application, 2001.
- SAMIER Henry, SANDOVAL Victor, *La recherche intelligente sur l'internet et l'intranet : outils et méthodes*, 2ème éd. revue et augmentée, Paris : Hermès Science Publications, 1999.
- SYS-COM, *Stratégie de test e-business*, Paris : Hermès Science Publications, 2001.

Liens Internet :

Choisir les mots-clés :

- *Optimisation de vos mots-clés pour le référencement [en ligne]*. Consulté le 20 août 2004. <http://www.webrankinfo.com/analyses/etudes/mots-cles.php>
- *Overture- Search Term Suggestion Tool [en ligne]*. Consulté le 20 août 2004. <http://inventory.overture.com/d/searchinventory/suggestion/>
- *Référencement, optimisation de référencement, positionnement [en ligne]*. Consulté le 20 août 2004. <http://www.123position.com/>
- SULLIVAN Danny. 08 avril 2002. *What People Search For - Most Popular Keywords [en ligne]*. Consulté le 20 août 2004. <http://searchenginewatch.com/facts/article.php/2156041>
- *Thesaurus.com [en ligne]*. Consulté le 20 août 2004. <http://thesaurus.reference.com/>

Le référencement :

- *Référencement et recherche d'info : tout sur le référencement de sites web avec Abondance [en ligne]*. Consulté le 20 août 2004. <http://www.abondance.com/>

- *Weborama.fr : classement quotidien des meilleurs sites francophones [en ligne]*. Consulté le 20 août 2004. <http://www.weborama.fr/>

Les moteurs de recherche :

- *IDF - les moteurs de recherche francophones - les moteurs de recherche francophones [en ligne]*. Consulté le 20 août 2004. <http://www.idf.net/mdr/>

- *INTERNET DIFFUSION. Guide complet des Moteurs de Recherche [en ligne]*. Consulté le 20 août 2004. <http://www.lesmoteursderecherche.com/>

Les annuaires :

- *Beaucoup! 2,000+ Search Engines, Indices and Directories [en ligne]*. Consulté le 20 août 2004. <http://www.beaucoup.com/>

- *Search engines directories metacrawler [en ligne]*. Consulté le 20 août 2004. <http://www.enfin.com/>

Web invisible :

- *BIBLIOTHEQUE NATIONALE DE FRANCE, DIRECTION DES COLLECTIONS. 01 juillet 2004. Les Signets de la Bibliothèques Nationale de France [en ligne]*. Consulté le 20 août 2004. <http://signets.bnf.fr/>

Étude de l'audience :

- *Bienvenue sur eStat - la mesure d'audience professionnelle fiable et simple [en ligne]*. Consulté le 20 août 2004. <http://www.estat.com/>

- *MEDIAMETRIE. Médiamétrie [en ligne]*. Consulté le 20 août 2004. <http://www.mediametrie.fr/web/>

- *Nedstat we're on your site [en ligne]*. Consulté le 20 août 2004. http://www.nedstat.com/fr/f50012_index.htm

- *Super Compteur - Outils d'analyse de trafic - webmasters professionnels [en ligne]*. Consulté le 20 août 2004. <http://supercompteur.francite.com/fr/index.asp>

- *Teoma - Search with Authority [en ligne]*. Consulté le 20 août 2004. <http://www.directhit.com/>

- Xiti : *Mesure d'audience professionnelle et analyse de R.O.I. AT INTERNET [en ligne]*. Consulté le 20 août 2004. <http://www.xiti.com/>

Autres méthodes de promotion :

- *Annuaire liste de diffusion, mailing liste, newsletter [en ligne]*. Consulté le 20 août 2004. <http://www.francopholiste.com/>

- *Conférence de Presse Virtuelle : Nous diffusons votre actualité [en ligne]*. Consulté le 20 août 2004. <http://www.conferencevirtuelle.com/>

- *Webring: [en ligne]*. Consulté le 20 août 2004. <http://dir.webring.com/rw>

Métadonnées et Dublin Core :

- ANDRIEU Olivier. Août 2002. *Abondance : les balises Meta servent-elles encore à quelque chose ? [en ligne]*. Consulté le 20 août 2004. <http://docs.abondance.com/question42.html>

- CHARLET Jean, LAUBLET Philippe, REYNAUD Chantal. *Web Sémantique : Présentation de l'Action Spécifique Web Sémantique (AS2W) STIC / CNRS [en ligne]*. Consulté le 20 août 2004. <http://www.lalic.paris4.sorbonne.fr/stic/presentation5.html>

- CHUNG Nicolas, CORMENIER Marie-Noëlle. Avril 2002. Normes et standards de métadonnées [en ligne]. Consulté le 20 août 2004. <http://www.educnet.education.fr/tech/normes/0402.htm>

- DUBLIN CORE METADAT INITIATIVE. VERCOUSTRE Anne-Marie (traduit par). 26 mars 2002. *Eléments de métadonnées du Dublin Core, Version 1.1: Description de Référence [en ligne]*. Consulté le 20 août 2004. <http://www-rocq1.inria.fr/~vercoust/METADATA/DC-fr.1.1.html>

- DUBLIN CORE METADATA INITIATIVE. 08 mars 2004. Dublin Core Metadata Initiative (DCMI) [en ligne]. Consulté le 15 août 2004. <http://dublincore.org/>

- DUBOST Karl. *RDF et les métadonnées - Slide list [en ligne]*. Consulté le 15 août 2004. [http://www.w3.org/2000/Talks/1019-rdf-poly/Karl Dubost](http://www.w3.org/2000/Talks/1019-rdf-poly/Karl%20Dubost)

- GOUVERNEMENT DU CANADA, COMMUNICATION CANADA, SECTEUR DES SERVICES DE COMMUNICATIONS, SITE DU CANADA. 05 mai 2001. *Guide des usagers sur les Métadonnées [en ligne]*. Consulté le 20 août 2004. http://198.103.99.147/publications/userman_f.html

- HAIGH Susan, Décembre 1999. *Flash Réseau n° 63 [en ligne]*. Consulté le 20 août 2004. <http://www.collectionscanada.ca/9/1/p1-262-f.html>

- HILLMANN Diane, TEASDALE Guy (traduit par). 15 janvier 2004. *Guide d'utilisation du Dublin Core [en ligne]*. Consulté le 15 août 2004. <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>
- JACQUET Christophe. 02 juin 2003. *Openweb.eu.org - Métadonnées et Dublin Core [en ligne]*. Consulté le 15 août 2004. http://www.openweb.eu.org/articles/dublin_core/
- MILLER Eric, SWICK Ralph, BRICKLEY Dan. 17 août 2004. *Resource Description Framework (RDF) / W3C Semantic Web Activity [en ligne]*. Consulté le 20 août 2004. <http://www.w3.org/RDF/>
- MINISTERE DE LA JEUNESSE, DE L'EDUCATION NATIONALE ET DE LA RECHERCHE. 14 août 2003. *Métadonnées, normes et standards : dossier Educnet [en ligne]*. Consulté le 20 août 2004. <http://www.educnet.education.fr/dossier/metadata/default.htm> 2003-08-14.
- MUTU-XML. 23 août 2004. *Dublin-Core [en ligne]*. Consulté le 20 août 2004. <http://www.mutu-xml.org/xml-base/shared/KEY-DUBLINCORE.html>
- PECCATTE Patrick. 10 mars 04. *Métadonnées: une initiation Dublin Core, IPTC, EXIF, RDF, XMP, etc. [en ligne]*. Consulté le 15 août 2004. <http://peccatte.karefil.com/software/Metadata.htm>

Glossaire

- **AICC** : (Aviation Industry Computer-based training Committee), créé en 1988, représente les aviateurs américains, avec une extension à d'autres secteurs d'activité grâce à son statut de pionnier. Ce comité s'intéresse à l'e-learning et sa réflexion sur les métadonnées est récente.

- **ARIADNE** : (Alliance of Remote Instructional Authoring and Distribution Networks for Europe), créé en 1996, rassemble des acteurs du secteur public et universitaire européen et suisse. Cette alliance est axée sur la FOAD (Formation Ouverte A Distance). Elle part des métadonnées du Dublin Core en les étendant.

- **ASCII** : La mémoire de l'ordinateur conserve toutes les données sous forme numérique. Il n'existe pas de méthode pour stocker directement les caractères. Chaque caractère possède donc son équivalent en code numérique : c'est le code ASCII (American Standard Code for Information Interchange - " Code Américain Standard pour l'Echange d'Informations"). Le code ASCII de base représentait les caractères sur 7 bits (c'est-à-dire 128 caractères possibles, de 0 à 127).

- **ASP** : Sigle de Active Server Pages. Il s'agit d'une technologie développée par Microsoft et liée aux scripts JavaScript, VBScript ou PerlScript incorporés dans des pages HTML. Au contraire du traitement classique - ces scripts sont interprétés sur l'ordinateur client par son navigateur Web -, ASP assure, sur le serveur Web, la compilation des scripts et leur conversion en code HTML au moment même de la demande de page. Cela permet d'inclure au vol des contenus provenant de bases de données du serveur, donc de proposer des pages Web " dynamiques ".

- **CGI**: (Anglais: Common Gateway Interface) une interface d'exécution des programmes sur un serveur.

- **DC** : Dublin Core, créé en 1995, rassemble des bibliothécaires et des documentalistes au niveau international, il définit des métadonnées génériques et développe des outils pour implémenter des métadonnées dans les ressources. Plus récemment, des groupes de travail sectoriels ont été mis en place, en particulier le Dublin Core Education pour ce domaine.

- **DTD** : La norme XML définit ainsi une définition de document type appelée DTD (Document Type Definition), c'est-à-dire une grammaire permettant de vérifier la conformité du document XML. La norme XML n'impose pas l'utilisation d'une DTD pour un document XML, mais elle impose par contre le respect exact des règles de base de la norme XML.

- **Flash** : Flash, de Macromedia, est le logiciel d'animation le plus utilisé pour compléter les pages Web avec des séquences graphiques animées. Le lecteur Flash permettant de visualiser ces animations est un module d'extension gratuit destiné aux navigateurs Web. C'est un tel standard de fait qu'il est souvent fourni avec le navigateur Web. Si vous n'en disposez pas et si vous consultez une page Web comportant des animations Flash, il vous est le plus souvent proposé de le télécharger directement.

- **IEEE** : (Institute of Electrical and Electronics Engineers), créé en 1998, basé aux Etats-Unis, rassemble des entreprises et des universités. Le sous-comité "technologies éducatives" de l'IEEE travaille sur la modélisation de tout objet d'apprentissage, il développe des métadonnées pour décrire les objets d'apprentissage y compris non numériques, il standardise aussi des vocabulaires (valeurs de métadonnées), les travaux de l'IEEE sont quasiment finalisés. Cet institut est à l'origine du LOM. Si l'on souhaite faire de la veille sur les métadonnées et le web sémantique, l'IEEE est l'institut dont il faut suivre les travaux.

- **IETF** : L'Internet Engineering Task Force est un groupe informel et auto organisé dont les membres contribuent à l'ingénierie et à l'évolution des technologies de l'Internet. C'est la principale structure engagée dans le développement des spécifications pour les nouveaux standards de l'Internet. L'IETF est atypique dans la mesure où elle est constituée d'une série d'événements, sans cadre statutaire ni conseil d'administration, sans membres ni adhésion.

- **Imagemaps** : Il est possible d'utiliser toute image comme une image cliquable (Image Map). On doit alors définir dans l'image des zones qui correspondront chacune à un URL (United Resource Locator) différent. Lorsque le lecteur cliquera sur ces zones dans l'image, le fureteur l'amènera à une ressource WWW locale où située sur un autre serveur.

- **IMS** : (Instructional Management Systems), créé en 1997, rassemble des universités et des entreprises américaines, est centré sur la FOAD et part des métadonnées du DC en les étendant.

- **Interopérabilité** : L'interopérabilité est le fait que plusieurs systèmes, qu'ils soient identiques ou radicalement différents, puissent communiquer sans ambiguïté. C'est une notion absolument cruciale pour le réseau téléphonique mondial et Internet, qui par essence sont des réseaux hétérogènes, où se croisent des matériels divers et variés, faisant tourner une panoplie encore plus vaste de logiciels.

- **ISBN** : L'ISBN (International Standard Book Number) est un numéro international normalisé qui permet d'identifier le titre d'un livre. Le numéro ISBN comporte 10 chiffres, répartis en 4 segments. Le premier identifie la zone linguistique (pour les pays francophones, c'est le 2) ; le deuxième indique l'éditeur (pour les publications officielles de l'administration centrale, c'est le 11) ; le troisième est le numéro d'ordre dans la production de l'éditeur : le quatrième sert de clé de contrôle.

- **ISO** : L'ISO est un réseau d'instituts nationaux de normalisation de 146 pays, selon le principe d'un membre par pays, dont le Secrétariat central, situé à Genève, Suisse, assure la coordination d'ensemble.

- **Java** : Java est un langage de programmation de haut niveau conçu par Sun Microsystems. Créé sous le nom " Oak ", il a été rebaptisé et aménagé pour servir plus particulièrement dans l'environnement du Web en 1995. Il s'agit d'un langage orienté objet inspiré du langage C++ mais simplifié afin d'en faciliter l'utilisation et la mise au point des programmes. Il est " semi-compilé ", c'est-à-dire que le code source d'un programme Java est converti en un code exécutable intermédiaire, le *bytecode*, qui n'est pas du code machine directement exécutable par le processeur mais du code interprété par un interpréteur. De petites applications Java, appelées applets, peuvent être automatiquement chargées sur le Web et exécutées par les navigateurs Web compatibles Java. La famille Java comporte également le JavaScript , le JDK , les JavaBeans et JDBC .

- **LOM (Learning Object Model)** : le LOM reprend toutes les métadonnées du Dublin Core en les détaillant davantage : structure de l'objet (structure, niveau d'agrégation), gestion de versions (version, état), gestion plus fine des contributeurs (rôle), description fine des métadonnées utilisées, description de l'implémentation technique au-delà du format : taille, localisation, prérequis techniques, gestion plus fine des droits (coûts...) et des relations entre objets, gestion de commentaires (annotations), gestion plus fine du sujet (par mots-clés ou classification). Pour la partie éducation, le LOM décrit les caractéristiques pédagogiques essentielles de l'objet concerné : type et niveau d'interactivité, type de ressource : exercice, QCM, examen, diagramme, figure, etc. « densité sémantique », type d'utilisateur : prof, élève, contexte (primaire, secondaire, supérieur, formation continue), classe d'âge typique, difficulté, durée, description, langue. Le document LOM va probablement être repris par l'ISO.

- **NCSA** : Le National Centre for Supercomputing Applications (ou littéralement Centre national pour les applications des super-ordinateurs) est un centre américain de recherche et d'exploitation des super-ordinateurs.

- **NISO** : NISO, l'organisme de normalisation national de l'information, une association sans but lucratif accréditée par le American National Standards Institute (norme ANSI), identifie, développe, maintient, et édite des normes techniques pour contrôler l'information dans un environnement numérique. Les normes de NISO appliquent la gamme de technologies traditionnelles et nouvelles entièrement des besoins information-connexes, y compris la récupération, re-purposing, le stockage, le metadata, et la conservation.

- **OCLC** : Fondé en 1967, l'Online Computer Library Centre est un organisme de recherche à but non lucratif qui offre des services aux bibliothèques afin qu'elles puissent accroître leur accès à l'information mondiale tout en réduisant leurs coûts. Plus de 50,000 bibliothèques dans 84 pays et territoires à travers le monde utilisent les services d'OCLC pour localiser, acquérir, cataloguer, prêter et préserver le matériel de bibliothèque.

- **Ontologie** : Organisation hiérarchique de la connaissance sur un ensemble d'objets par leur regroupement en sous-catégories suivant leurs caractéristiques essentielles.

- **RDF** : (Resource Description Framework) langage XML de métadonnées qui permet de représenter des relations entre ressources, sous la forme de graphes associant des ensembles de « triplets » élémentaires. RDF constitue la base technologique de la vision du Web sémantique de Tim Berners-Lee.

- **Robot** : Le moteur de recherche est alimenté automatiquement par un robot logiciel. Le robot a pour mission de parcourir en permanence l'Internet à la recherche de nouvelles pages qu'il suit de lien en lien, afin de nourrir et d'enrichir sa base de données. Chaque moteur est unique et possède son propre algorithme conçu pour implémenter le classement des listes de résultats.

- **SCORM** : (Sharable Content Object Reference Model), créé en 1997, rassemble des acteurs du ministère de la défense aux États-unis et est centré entièrement sur la FOAD, ce modèle reprend l'IMS et l'AICC.

- **Script** : ensemble de commandes capables d'automatiser certaines tâches d'un programme de communication.

- **Spamdexing** : utilisation de techniques « frauduleuses » pour essayer de placer une page le plus haut possible dans les résultats des moteurs de recherche.

- **Topic Maps** : L'objet d'une *Topic Map* est de transporter dans une couche indépendante des ressources la connaissance qui s'y rapporte. . Une *Topic Map* réunit les sujets traités par les ressources ainsi que les relations qui existent entre eux. Cela d'une façon totalement indépendante des méthodes de mise en œuvre.

- **URI** : Un URI, sigle Uniform Resource Identifier en anglais, littéralement « identifiant uniforme de ressource », est une chaîne de caractères respectant le format défini par une norme d'Internet et servant à identifier des entités éventuellement accessible sur Internet. Les URI forment l'ensemble de tous les noms et adresses qui, sous forme de chaînes de caractères, réfèrent à des ressources sur le Web.

- **URL** : Sigle de Uniform Resource Locator (localisateur uniformisé de ressource). Sur Internet, différents types de ressources sont disponibles : pages Web, fichiers à télécharger, boîtes aux lettres électroniques, forums de discussion, etc. Pour faciliter l'accès à ces ressources, Tim Berners-Lee, le créateur du Web , a proposé en 1994 sous cette dénomination d'URL une syntaxe uniformisée permettant, en une seule ligne de commande, d'accéder à tout type de ressource. Le format générique des URL est le suivant : type de service : infos d'accès dépendant du type de service. Ainsi, qu'il s'agisse du service Web (http), FTP, e-mail (mailto), Usenet (news ou nntp) ou d'un autre type de service, l'utilisateur n'a pas besoin d'utiliser plusieurs commandes (demande de connexion au serveur, commande spéciale, etc.) mais une seule pour accéder à une ressource. Pour bon nombre de services, les informations d'accès se composent du nom DNS ou de l'adresse IP du serveur sur lequel se trouve la ressource et du chemin d'accès à cette ressource. Elles peuvent aussi contenir un nom d'utilisateur, un mot de passe et un numéro de port logique si nécessaire.

- **W3C** : (World Wide Web consortium) Consortium international dont le but est de promouvoir l'évolutivité du Web et de garantir son interopérabilité.

- **XML** : eXtensible Mark-up Language (Langage par balises extensible). Langage de description de document, successeur annoncé de SGML et de HTML. XML est le nouveau langage de description de documents normalisé par le W3C, notamment pour la création de pages Web.

Annexes

Annexe 1 : Compte-rendu de l'intervention de Laurent Trémel lors de la journée professionnelle du 27/11/2003 à l'IUFM d'Arras : « Télévision, jeux vidéo : une alternative aux livres ? » par Jonathan Marie et Julien Siméoni.

Source : MARIE Jonathan, SIMEONI Julien, *Faut-il stigmatiser les jeux vidéo ?*, Lille : UFR IDIST, 2004. http://www.univ-lille3.fr/jeunet/jpro/sommaire/fr_som.htm.

Faut-il stigmatiser les jeux vidéo ?

Les jeux vidéo peuvent-ils, au même titre que le livre, contribuer à la formation de la personnalité des adolescents ? Si oui, pour développer quels types d'imaginaires ? Telle était la question posée à Laurent Trémel, sociologue et auteur de *Jeux de rôles, jeux vidéos, multimédia : les faiseurs de mondes* paru chez PUF en 2001. Son exposé s'est dans un premier temps appuyé sur les recherches menées sur le sujet, et en particulier sur les travaux de Pierre Bruno parus (1993) et de Tony Fortin (2004).

Le jeu de rôles, une pratique de contestation

Les jeux de rôles ont comme caractéristique des univers " dérangeants ", fortement hiérarchisés, reposant sur de fortes inégalités entre les êtres. Le propre du personnage est de " se distinguer du commun ". Les joueurs sont majoritairement des enfants de cadres et souvent en situation de rejet de l'école : ils sont sceptiques face aux réformes sociales, cyniques face à l'avenir qui les attend. Les univers qu'on leur propose se diversifient et passent progressivement de l'heroic fantasy au cyberpunk, monde dans lequel les personnages se battent contre une politique ultra capitaliste, où les média sont tout puissants. Jouer est un moyen de s'opposer à une société dont on n'approuve pas les valeurs, de se forger une identité, même fictive, dans l'objectif d'accroître sa grandeur au sein du microcosme. Les jeux de rôles proposent une alternative, un système différent dans lequel il est possible d'évoluer et d'acquérir une importance. Ils permettent une évolution virtuelle par étapes, qui compense la stagnation des joueurs dans la vie sociale.

Les jeux vidéo, pas si violents qu'on le dit

Pour Laurent Tremel, les jeux vidéos sont les héritiers des jeux de rôles. Ils ont cependant connu un destin différent dans les média. Certains jeux ont fait l'objet de critiques tandis que d'autres ont été encensés en raison de leur caractère pédagogique. Les pratiquants des jeux vidéo sont pour la plupart des enfants de cadres dont les parents adhèrent aux valeurs de l'informatique et du multimédia. Ils se posent moins de questions sur leur environnement social que les adeptes des jeux de rôles et adhèrent au discours socialement accepté. On ne retrouve pas chez eux le désir de se marginaliser par le jeu.

Il existe cependant une minorité plus contestataire : les " hardcore gamers " au discours moins politiquement correct. Ce terme s'applique aux joueurs acharnés, qui jouent énormément, au détriment de leur vie sociale. Ces joueurs sont plus critiques et trouvent dans les jeux une image d'eux-mêmes et de la société qui leur convient mieux que celle que leur renvoie la réalité.

Pour ce qui concerne la question de la violence, Laurent Tremel va à l'encontre des discours alarmistes. Cette idée, qu'il pas le premier à défendre, s'appuie sur le fait que le joueur doit acquérir les codes sur lesquels reposent les jeux. Il prend par conséquent du recul par rapport

à la violence qu'il y rencontre. Il comprend dès lors le caractère fictionnel du jeu et le considère un simple outil de distraction détaché de la réalité. En revanche le joueur qui n'a pas intégré ces codes risque de ne pas avoir cette attitude distanciée. Les jeux vidéo nécessitent donc recul et discernement pour éviter les pièges de la violence. A l'opposé, Laurent Trémel critique les "discours positifs" sur certains jeux vidéo, ces discours étant parfois plus démagogiques que scientifiques.

De bons jeux vidéo ?

Dans son étude, Laurent Trémel s'est intéressé particulièrement aux " bons jeux vidéo sur PC ", considérés généralement comme " intelligents et pédagogiques ". Sous des dehors de fiction historique, ces jeux cachent une conception néo-libérale du monde. Il prend ici l'exemple de Civilization, qui véhicule en fait un discours fortement proaméricain. Le joueur qui choisit de jouer avec les Etats-Unis se voit en effet doté d'une armée bien mieux équipée que ses adversaires. Celui qui choisit les Zoulous ou les Mongols se voit attribuer des guerriers féroces qui ne disposent pas des moyens des pays développés, ce qui lui laisse peu de chances de gagner. Quant aux Français, ils se trouvent constamment envahis et contraints à recourir à des aides extérieures pour se libérer.

Un autre exemple est celui de Fire Department, qui a fait l'objet d'un travail de Tony Fortin, responsable du site "Planetjeux", un jeu de simulation dans lequel le joueur dirige une équipe de pompiers chargée d'éteindre une maison de star en feu. Les pompiers reçoivent comme consigne de sauver en priorité la star, puis sa voiture et en tout dernier lieu le jardinier ! Dans le même esprit, il est plus urgent de sauver l'argent que les clients dans une banque en flamme.

La conclusion est difficile !

De par leur utilisation par le plus grand nombre, les jeux vidéo perdent ainsi la dimension artisanale et underground qui caractérisait les jeux de rôles. Ils ne sont pas conçus pour susciter l'esprit critique, mais pour provoquer l'adhésion aux valeurs de la société moderne. Les statistiques montrent par ailleurs que les joueurs vidéo, qui sont plutôt issus de milieux favorisés, sont aussi pour la plupart de meilleurs lecteurs que la moyenne, ce qui ne facilite pas la réflexion sur la pratique. Le recul remarqué peut très bien provenir d'une culture livresque.

Notre avis pour terminer

En bref cette intervention fut globalement intéressante, la critique faite des jeux dits " intelligents " passionnante. Pour les joueurs de jeux vidéo que nous sommes, le fait que l'étude n'ait porté que sur ces jeux pratiqués par un public cultivé représente une déception. Il existe d'autres jeux réputés intelligents qui n'ont pas été évoqués, et surtout une multitude d'autres auxquels on n'attribue pas cette qualité et dont nous n'avons rien su. Laurent Trémel nous répond qu'un travail scientifique repose sur un corpus délimité, au contraire d'une émission de télévision. Dont acte !

Trémel, Laurent. *Jeux de rôles, jeux vidéo, multimédia : les faiseurs de mondes*. Paris : PUF, 2001. 309 p. (Sociologie d'aujourd'hui).

Le grand jeu : débats autour de quelques avatars médiatiques, sous la direction de Nicolas Santolaria et de Laurent Trémel. Paris : PUF, 2004. 199 p. (Sociologie d'aujourd'hui). Cet ouvrage comporte un article de Tony Fortin.

Annexe 2 : Contexte de la recherche d'informations dans un système documentaire classique et sur l'Internet.

Source : LEFEVRE Philippe, *La recherche d'informations : du texte intégral au thésaurus*, Paris : Hermès Science Publications, 2000, p.205

Systèmes documentaires classiques	Internet
Taille de la base limitée – croissance linéaire avec le temps	Quantité d'informations : des millions de gigaoctets – croissance exponentielle
Informations secondaires	Informations primaires mélangées à des informations secondaires
Bases internes, d'où responsabilité de constitution, de mise à jour et d'indexation de la base	Bases externes sur lesquelles aucune action n'est possible, en particulier pour l'indexation
Homogénéité des sources et des informations – la base recouvre un ou plusieurs domaines bien circonscrits, avec un approche précise : scientifique, commerciale ou autre	Hétérogénéité des sources et des informations – on trouve tout sur les réseaux – plus de domaine restreint, toutes les approches : scientifique, vulgarisation, commerciale, etc.
Format unique de représentation des informations.	Formats multiples et en évolution incessante
Bases structurées (champs de catalogage renseignés)	Sites Web pas du tous structurés – utilisation minimale des balises méta
Existence d'un langage d'indexation (thésaurus) adapté au domaine et à la base	Pas de langage d'indexation : il n'existe pas de thésaurus couvrant l'ensemble des domaines de connaissance – trop d'acteurs différents interviennent
Indexation en utilisant le thésaurus adapté	Indexation des moteurs de recherche sur le Web de très bas niveau : niveau 0 ou 0+ selon notre classification
Indexation de la totalité du corpus	Seulement 15 % du Web indexé par les moteurs les plus puissants – 2 formes de sites Web invisibles
Informations entrées dans la base vérifiées – sources connues	Information non validées – nécessité de qualifier les informations trouvées
Monolinguisme, au moins des index ou descriptions	Multilinguisme
Outil d'interrogation unique, adapté à la base et au langage utilisé	L'interrogation dépend du moteur utilisé – mélange des modes de recherche : répertoires, mots, navigation.
Interrogation réalisée par un professionnel qui connaît le langage d'indexation et de requêtes	L'accès n'est plus réservé aux professionnels de la documentation : tout le monde se connecte
Alimentation réalisée par un documentaliste ayant des soucis de cohérence des informations	Alimentation réalisée par un <i>Webmaster</i> informaticien ayant des soucis de présentation, de diffusion et d'impact des informations

Annexe 3 : Utilisation de la syntaxe HTML avec le Dublin Core

Source : HILLMANN Diane. TEASDALE Guy (traduit par).15 janvier 2004. *Guide d'utilisation du Dublin Core [en ligne]*. Consulté le 15 août 2004. <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>

Chaque définition d'élément descriptif a un attribut NAME et un attribut CONTENT comme dans:

```
<META NAME="DC.Creator" CONTENT="Browning, Elizabeth">
```

N'importe quel élément de métadonnées peut être omis ou répété. Lorsqu'on répète un élément, la pratique recommandée est de lister chaque définition d'élément séparément comme dans:

```
<META NAME="DC.Creator" CONTENT="Marx, Karl">  
<META NAME="DC.Creator" CONTENT="Engels, Friedrich">
```

Toutefois, il est aussi valide d'exprimer des éléments répétés en utilisant un seul attribut NAME avec plusieurs valeurs pour l'attribut CONTENT, séparées par des points virgules comme dans:

```
<META NAME="DC.Creator" CONTENT="Marx, Karl; Engels, Friedrich">
```

La convention proposée ici pour inclure des métadonnées en HTML s'accorde avec une autre convention pour identifier et grouper différents schémas de métadonnées en HTML. Elle repose sur l'utilisation d'un préfixe pour indiquer que les éléments utilisés proviennent du Dublin Core ou d'un autre schéma de métadonnées. Pour faciliter la lecture, le préfixe "DC" devrait être écrit en lettres majuscules et les noms d'éléments devraient s'écrire avec une majuscule initiale. Par exemple:

```
META NAME="DC.Title"  
META NAME="DC.Creator"
```

ET NON PAS

DC.CREATOR ou dc.CREATOR ou DC.creator

Si des caractères non-ASCII sont requis, utiliser les mêmes conventions que dans le corps du document. Par exemple:

```
<META NAME="DC.Title" CONTENT="Les biscuits &agrave; la banane">
```

Annexe 4 : Champs de métadonnées du Dublin Core

Source : JACQUET Christophe. 02 juin 2003. *Openweb.eu.org - Métadonnées et Dublin Core [en ligne]*. Consulté le 15 août 2004. http://www.openweb.eu.org/articles/dublin_core/

Liste des éléments du Dublin Core	
Élément	Description et liste des raffinements
title	<i>Titre du document</i> : il s'agit a priori du titre principal du document. Pour indiquer un autre type, on peut utiliser le raffinement suivant : <i>alternative</i> : alternative pour le titre, par exemple une abréviation ou une traduction.
creator	<i>Créateur du document</i> : nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document.
subject	<i>Sujet et mots-clefs</i> : mots-clefs, phrases de résumé, ou codes de classement. Il est préférable d'utiliser des mots-clefs choisis dans le cadre d'une politique de classement. Par exemple, il est recommandé d'utiliser les codages de la bibliothèque du congrès (LCSH et LCC), le vocabulaire médical (MESH), ou les notations décimales des bibliothécaires (DDC et UDC).
description	<i>Description du document</i> : résumé, table des matières, ou texte libre. Le type de description peut être précisé à l'aide des raffinements suivants : <i>tableOfContents</i> : table des matières ; <i>abstract</i> : résumé.
publisher	<i>Editeur du document</i> : nom de la personne, de l'organisation ou du service à l'origine de la publication du document.
contributor	<i>Contributeur au document</i> : nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document.
date	<i>Date d'un évènement dans le cycle de vie du document</i> : il peut s'agir par exemple de la date de création ou de la date de mise à disposition. Il est recommandé de spécifier la date au format W3CDTF (AAAA-MM-JJ). Pour préciser de quelle date il s'agit, on utilise les raffinements suivants : <i>created</i> : date de création ; <i>valid</i> : date ou période de validité ; <i>available</i> : date ou période de mise à disposition ; <i>issued</i> : date de publication ; <i>modified</i> : date de modification ; <i>dateAccepted</i> : date d'acceptation (par exemple, acceptation d'une thèse par une université, d'un article par un journal, etc.) ; <i>dateCopyrighted</i> : date du copyright ; <i>dateSubmitted</i> : date où le document a été soumis (par exemple, soumis à un comité de lecture s'il s'agit d'un article).
type	<i>Nature ou genre du contenu</i> : grandes catégories de document. Il est recommandé d'utiliser des termes clairement définis au sein de son organisation. Par exemple, le Dublin Core définit quelques types dans le vocabulaire DCMITypes.
format	<i>Format du document</i> : format physique ou électronique du document. Par exemple, type de média ou dimensions (taille, durée). On peut spécifier le matériel et le logiciel nécessaires pour accéder au document. Il est recommandé d'utiliser des termes clairement définis, par exemple les types MIME ou IMT . Les raffinements suivants sont disponibles : <i>extent</i> : taille ou durée ; <i>medium</i> : support physique.
identifier	<i>Identificateur non ambigu</i> : il est recommandé d'utiliser un système de référencement précis, par exemple les URI ou les numéros ISBN.
source	<i>Ressource dont dérive le document</i> : le document peut découler en totalité ou en partie de la ressource en question. Il est recommandé d'utiliser une dénomination formelle des ressources, par exemple leur URI.

language	<i>Langue du document</i> : il est recommandé d'utiliser un code de langue conforme au format RFC3066.
relation	<p><i>Lien vers une ressource liée</i> : il est recommandé d'utiliser une dénomination formelle des ressources, par exemple leur URI. On précise le type de lien avec des raffinements :</p> <p><i>isVersionOf</i> : on a affaire à une nouvelle version, une modification ou une adaptation du document lié. Les changements concernent le contenu et pas seulement la forme ;</p> <p><i>hasVersion</i> : réciproque d'<i>isVersionOf</i>. Le document lié est une version modifiée du présent document ;</p> <p><i>isReplacedBy</i> : le présent document a été remplacé par le document lié ;</p> <p><i>replaces</i> : réciproque de <i>replaces</i>. Le présent document remplace le document lié ;</p> <p><i>isRequiredBy</i> : on a besoin du présent document pour interpréter correctement le document lié ;</p> <p><i>requires</i> : réciproque d'<i>isRequiredBy</i>. Le présent document a besoin du document lié pour être correctement présenté, transmis, ou pour assurer sa cohérence ;</p> <p><i>isPartOf</i> : le document est une partie (physique ou logique) d'un autre document ;</p> <p><i>hasPart</i> : réciproque d'<i>isPartOf</i> : le document inclut le document lié, physiquement ou logiquement ;</p> <p><i>isReferencedBy</i> : le document courant est référencé, cité, ou lié par le document indiqué ;</p> <p><i>references</i> : réciproque d'<i>isReferencedBy</i> : le document courant référence, cite ou pointe vers le document indiqué ;</p> <p><i>isFormatOf</i> : le présent document a le même fond que le document indiqué, mais présenté sous une forme différente ;</p> <p><i>hasFormat</i> : réciproque d'<i>isFormatOf</i> : le présent document possède une variante sous une forme différente ;</p> <p><i>conformsTo</i> : référence à un standard établi auquel se conforme le présent document.</p>
coverage	<p><i>Portée du document</i> : la portée inclut un domaine géographique, un laps de temps, ou une juridiction (nom d'une entité administrative). Il est recommandé d'utiliser des représentations normalisées de ces types de données. Le type de couverture peut être précisé :</p> <p><i>spatial</i> : couverture spatiale. On peut utiliser les codages Point (point géographique), ISO3166 (codes de pays à deux lettres), Box (régions géographiques), ou TGN (dictionnaire de noms de lieux) ;</p> <p><i>temporal</i> : couverture temporelle. On peut utiliser les codages Period (intervalle de temps) ou W3CDTF (dates).</p>
rights	<i>Droits relatifs à la ressource</i> : permet de donner des informations sur le statut des droits du document, par exemple la présence d'un copyright, ou un lien vers le détenteur des droits. L'absence de cet élément ne présume pas que le document est libre de droits.
audience	<p><i>Audience du document</i> : l'audience représente le groupe de personnes à qui le document est destiné. L'audience est déterminée par l'auteur, le publicateur, ou un tiers. On peut utiliser les raffinements suivants :</p> <p><i>mediator</i> : entité qui sert d'intermédiaire pour l'accès au document ;</p> <p><i>educationLevel</i> : position du niveau de l'audience par rapport à un contexte d'éducation ou de formation.</p>

Annexe 5 : Formats de données conseillés par le Dublin Core

Source : JACQUET Christophe. 02 juin 2003. *Openweb.eu.org – Métadonnées et Dublin Core [en ligne]*. Consulté le 15 août 2004. http://www.openweb.eu.org/articles/dublin_core/

Nom formel	Description	Éléments concernés
LCSH	En-tête de la bibliothèque du Congrès (Library of Congress Subject Heading).	subject
MESH	Vocabulaire médical MeSH (Medical Subject Headings).	Subject
DDC	Notation décimale de bibliothèque Dewey Decimal Classification.	subject
LCC	Classification de la bibliothèque du Congrès (Library of Congress Classification).	subject
UDC	Notation décimale de bibliothèque Universal Decimal Classification.	subject
DCMIType	Type du document parmi une série de types définis par le Dublin Core : Collection (agrégation de documents formant un tout), <i>Dataset</i> (structure appropriée au traitement par ordinateur), <i>Event</i> (événement, par définition éphémère). Par exemple : exposition, conférence, bataille, procès, mariage, etc.), <i>Image</i> (photographie, peinture, dessin, animation, film, diagramme, carte, notation musicale, etc.), <i>InteractiveResource</i> (objet qui demande une interaction de l'utilisateur, par exemple une page Web, une applet, un service de chat, etc.), <i>Service</i> (entité qui rend des services à l'utilisateur, comme un service de photocopie ou un serveur Web), <i>Software</i> (logiciel d'ordinateur), <i>Sound</i> (son, par exemple voix ou musique), <i>Text</i> (texte au sens large : livre, magazine, journal, poème, message d'une liste de diffusion, etc.), <i>PhysicalObject</i> (tout objet physique : ordinateur, sculpture, bâtiment, etc.)	type
IMT	Internet Media Types, nomenclature de types de données très proche MIME (exemple : image/jpeg)	format
ISO639-2	Code de langue à trois lettres ISO-639-2 (du type fre).	language
RFC1766	Code de langue RFC 1766. Il est composé d'un code de langue sur deux lettres, éventuellement suivi d'un suffixe de pays pour indiquer la variante de la langue (par exemple, fr-CA pour la Français du Canada, fr-FR pour le Français de France, en pour l'Anglais en général).	language
RFC3066	Code de langue RFC 3066, nouvelle version de RFC1766 (le code de langue peut être sur trois lettres).	language
URI	Un URI, selon la RFC 2396.	identifier, source, relation
Point	Point de l'espace repéré par ses coordonnées géographiques.	coverage.spatial
ISO3166	Codes ISO-3166 de représentation des noms de pays sous forme de chaîne de deux caractères (exemples : FR, DE).	coverage.spatial
Box	Représentation des régions géographiques du Dublin Core.	coverage.spatial
TGN	Noms issus du Getty Thesaurus of Geographic Names.	coverage.spatial
Period	Représentation des intervalles de temps du Dublin Core.	coverage.temporal, date
W3CDTF	Codage des dates et heures du W3C. Pour référencer un jour, on utilise le format AAAA-MM-JJ.	coverage.temporal, date

Annexe 6 : Composition d'une page HTML

Source : ESKENAZI Jean-Pierre F., *Référencement : comment référencer son site Web*, Paris : Webedition, 1999, p.275-276.

Une page HTML, à l'instar des publications professionnelles, est structurée. Elle est généralement constituée en trois parties, placées entre des balises de début et de fin :

- Le format de la page,
- L'en-tête de page,
- Le corps de page.

- Le format de la page est placé entre les balises `<HTML>` et `</HTML>`, ce qui indique au navigateur que votre page est au format HTML. Les deux parties qui suivent sont placées entre ces balises.

- L'en-tête de page est placée entre les balises `<HEAD>` et `</HEAD>`, et informe du contenu du document. Invisible lors de la consultation d'une page ; l'en-tête peut être affichée en fichier source.

La section d'en-tête peut contenir six commandes différentes :

TITLE : Titre du document (obligatoire)

META : Incorporation d'informations supplémentaires

LINK : Liens avec d'autres documents

BASE : Origine du document

ISINDEX : Autorise la recherche sur le document

NEXTID : Identificateur de document unique

Dans l'objectif du référencement, l'en-tête doit contenir au minimum les balises de titre de la page, de descriptif, ainsi que la liste des mots-clés, et ce sur chaque page soumise à l'indexation.

- Le corps de page, partie du document affichée par le navigateur, est placé entre les balises `<BODY>` et `</BODY>`.



Annexe 7 : Valeur ASCII des caractères



Source : ESKENAZI Jean-Pierre F., *Référencement : comment référencer son site Web*, Versailles : Webedition, 1999, p. 87-88.


Les caractères ASCII 0000 à 0008 et 0011 à 0031 et 0127 à 0160 ne sont pas utilisés.

	0009	Tabulation	0010	Retour à la ligne	0032	Espace					
0003	!	0034	"	0035	#	0036	\$	0037	%	0038	&
0039	'	0040	(0041)	0042	*	0043	+	0044	,
0045	-	0046	.	0047	/	0048	0	0049	1	0050	2
0051	3	0052	4	0053	5	0054	6	0055	7	0056	8
0057	9	0058	:	0059	;	0060	<	0061	=	0062	>
0063	?	0064	@	0065	A	0066	B	0067	C	0068	D
0069	E	0070	F	0071	G	0072	H	0073	I	0074	J
0075	K	0076	L	0077	M	0078	N	0079	O	0080	P
0081	Q	0082	R	0083	S	0084	T	0085	U	0086	V
0087	W	0088	X	0089	Y	0090	Z	0091	[0092	\
0093]	0094	^	0095	_	0096	`	0097	a	0098	b
0099	c	0100	d	0101	e	0102	f	0103	g	0104	h
0105	i	0106	j	0107	k	0108	l	0109	m	0110	n
0111	o	0112	p	0113	q	0114	r	0115	s	0116	t
0117	u	0118	v	0119	w	0120	x	0121	y	0122	z
0123	{	0124		0125	}	0126	~				

Les caractères ASCII 0127 à 0160 ne sont pas utilisés ou sont réservés en localisation selon certains alphabets ou caractères étendus, comme :


 0145 pour ' suivi de  0146 pour ' ;


 0147 pour " suivi de  0148 pour " ;



 0160 pour ;



etc.



Enfin les caractères ASCII suivants, 0161 à 0255, sont représentatifs des caractères accentués ou latins spécifiques, comme :

 0181 pour μ ;

 0182 pour ¶ (signe de fin de paragraphe) ;


 0203 pour È ou  0235 pour è ;





 0140 pour Œ ou  0156 pour œ ;

 0169 pour © (copyright) ou  0174 pour ® (marque déposée) ;

etc.

Remarque

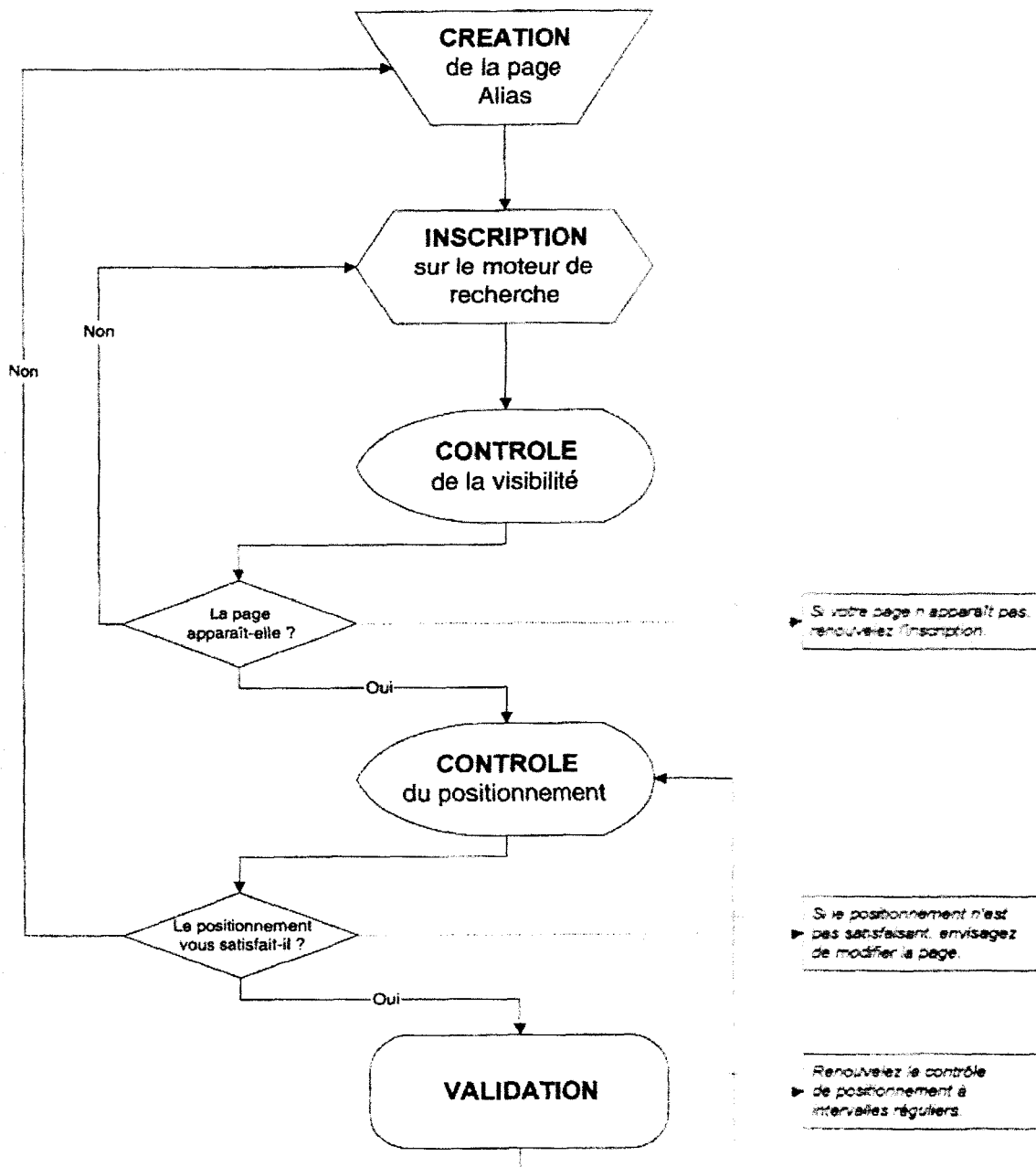
A partir de Windows 98, le caractère monétaire de l'Euro  est intégré à la table ASCII :

- accès direct par    (maintien en appui des trois touches) ;
- caractère ASCII  0128 .
- ou par une touche spécifique sur les claviers récents.

Attention : ce caractère ne s'affichera sur un site que si les polices présentes sur le poste de l'internaute le reconnaissent.

Annexe 8 : Schéma d'inscription pour une page sur un moteur

Source : ESKENAZI Jean-Pierre F., *Référencement : comment référencer son site Web*, Paris : Webedition, 1999, p.105.



Annexe 9 : Code d'une page en HTML corrigée pour gérer au mieux les frames

Source : PUBLICITYWEB, *Référencer et promouvoir son site*, Paris : Micro Application, 2001, p. 110-111.

```
<html>
<head>

<TITLE> Auberge des mouflons, Val d'Isère, Alpes - France </TITLE>

<meta name="description" content="L'auberge des mouflons à Val d'Isère
➤ dans les Alpes Française en Savoie, offre une splendide vue
➤ panoramique sur la vallée et les montagnes. Chambres d'hôtes,
➤ demi-pension et restaurant.">

<meta name="keywords" content="auberge, restaurant, hôtel, logement,
➤ hôte, restauration, chambre, lit, nuitée, gîte, résidence, cuisine,
➤ manger, dormir, menu, montagne, montagnard, sommet, pic, neige,
➤ alpes, alpage, savoie, france, français, française, Auberge,
➤ Restaurant, Hôtel, Logement, Hôte, Restauration, Chambre, Lit,
➤ Nuitée, Gîte, Résidence, Cuisine, Manger, Dormir, Menu, Montagne,
➤ Montagnard, Sommet, Pic, Neige, Alpes, Alpage, Savoie, France,
➤ Français, Française, auberges, restaurants, hôtels, logements, hôtes,
➤ restaurations, chambres, lits, nuitées, gîtes, résidences, cuisines,
➤ manger, dormir, menus, montagnes, montagnards, sommets, pics, neiges,
➤ alpes, alpages, savoyard, France, français, françaises">

</head>

<frameset rows="64,*">
  <frame name="top" scrolling="no" noresize target="contenu"
  ➤ src="top.html">
  <frameset cols="150,*">
    <frame name="navigation" target="contenu" src="navigation.html">
    <frame name="contenu" src="contenu.html">
  </frameset>
</frameset>

<noframes>
<body>

  <h1>L'auberge
  des mouflons à Val D'Isère dans les Alpes Française en Savoie, offre
  ➤ une splendide vue panoramique sur la vallée et les montagnes.
  ➤ Chambres d'hôtes,
  demi-pension et restaurant.</h1>

  <p>Cette page utilise des frames, mais votre navigateur ne les prend pas
  en charge. Nous vous conseillons d'installer la dernière version de Netscape
  Navigator ou de Microsoft Internet Explorer.</p>

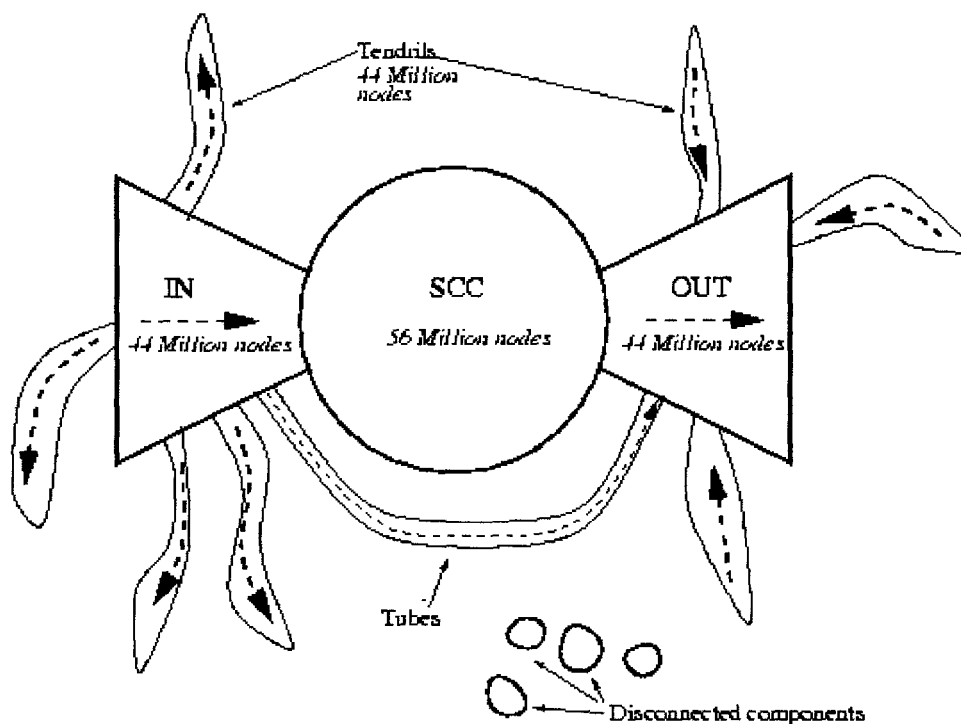
  <p>Cependant, pour vous permettre de découvrir ce que nous proposons,
  veuillez suivre les liens ci-dessous :</p>

  <p>
  - <a href="contenu.html">Accueil à l'Auberge</a><br>
  - <a href="chambres.html">Les chambres d'hôtes</a><br>
  - <a href="restaurant.html">Le restaurant</a><br>
  - <a href="activités.html">Les activités</a><br>
  - <a href="region.html">La région</a>
  </p>

</body>
</noframes>
</frameset>
```

Annexe 10 : Structure du Web en « nœud papillon »

Source : FOENIX-RIOU Béatrice, *Recherche et veille sur le Web visible et invisible : agents intelligents, annuaires sélectifs, interfaces des grands serveurs, portails thématiques*, Paris : Bases, Paris : Éditions technique et documentation, 2001, p. 12-13.



- *La partie centrale (SCC – strongly connected component – sur la figure) est constituée du « noyau ultra-connecté » et contient moins d'un tiers des pages Web (28%). La navigation y est aisée, car chacune des pages est reliée aux autres par des chemins de liens hyperterxte.*
Ce noyau compact constitue le cœur du réseau Internet. C'est lui qui permet de passer, par clics successifs, de n'importe quelle page du IN vers une page de OUT. Ce sont les pages du cœur que les robots des moteurs de recherche indexent en priorité et c'est à partir de leurs liens qu'ils explorent le Web.
- *La partie gauche (IN) contient les pages d'origine et représente environ un cinquième du réseau (21%).*
Ces pages offrent des liens vers le cœur du Web, mais l'inverse n'est pas vrai ; on trouve dans cette catégorie, par exemple, les pages de moindre intérêt pour la communauté des netsurfers (certaines pages personnelles...) ou les pages de création récente, qui n'ont pas été reconnues par leurs pairs et vers lesquelles ne pointent encore que peu de clients.
- *La partie droite (OUT) correspond aux pages de destination : elle représente également un cinquième du réseau. Ces pages sont accessibles depuis le cœur du Web, mais aucun retour n'est possible. On trouve dans cette catégorie, notamment, les sites commerciaux (sites d'entreprises, de commerce électronique...), vers lesquels pointent de nombreux liens, mais qui, eux, n'en proposent pas, ou seulement en interne.*
- *Une dernière zone, représentant également un cinquième du Web (Tendrils), est composée de pages non connectées au cœur du réseau.*
Ces pages sont accessibles depuis les pages d'origine et/ou donnent accès aux pages de destination.
- Enfin, près de 10% des pages Web sont totalement déconnectées des autres pages.

Résumé

De nos jours, la conception d'un site Internet ne peut se faire sans y inclure la préparation à l'indexation et au référencement. Ces deux opérations sont devenues indispensables pour qu'un site ne soit pas noyé dans la masse informationnelle et pour qu'il ait des chances de ressortir en résultat à des requêtes. Quelles sont les méthodes pour s'acquitter au mieux de ces tâches ? Quel est leur avenir dans un domaine où la compétition commerciale se taille de plus en plus la part du lion au détriment d'une précision documentaire ?

Abstract

Nowadays, more and more websites are created. It induces an increase of the information flow. Hence, it becomes hard to promote efficiently one's site among the mass. A good promotion must include a good indexing and a good referencing. Both processes need rigour. However, the Web is currently ruled by financial interests which require more competitiveness than accuracy. What is therefore the use of methods made to improve indexing and referencing?

Mots-clés : indexation, référencement, pages web, sites web, littérature de jeunesse, webmaistrance.

Keywords: indexing, referencing, web pages, web sites, youth literature, webmaster.