



Les tablettes...

d'Ourouk.

**Référencement et
visibilité d'un site
web**

Numéro 7

Septembre 2002

Les enjeux

Le développement exponentiel des sources d'information sur le web dans les dernières années (plusieurs milliards de pages web) et l'absence de règles d'organisation et de hiérarchisation de l'information ont généré des besoins nouveaux pour les utilisateurs du web, qu'ils soient consommateurs ou producteurs d'information. En effet, la difficulté majeure pour l'internaute qui recherche une information est d'être le plus souvent noyé sous des centaines de réponses à sa requête, réponses dont il ne consultera le plus souvent que les vingt premières. Par ailleurs, le producteur d'information (entreprise, administration, organisme, association...) souhaite que l'information qui a été mise en ligne soit accessible facilement et rapidement par son public, et au-delà, dans un environnement où l'information est surabondante mais pas toujours de qualité.

Ces besoins spécifiques des acteurs du web mettent en évidence deux processus fondamentaux pour aboutir à une présence efficace de l'information sur le web à savoir le **Positionnement** et le **Référencement** sur les outils de recherche.

Le *positionnement* est défini comme le classement d'une page web dans l'ensemble des liens proposés à l'internaute à la suite de sa requête sur un moteur de recherche. La notion de positionnement est primordiale en terme de visibilité sur le web dans la mesure où l'ensemble des études montrent que la majorité des internautes explorent les vingt premières réponses proposées (en général les deux premières pages de réponses fournies par le moteur de recherche). L'objectif en terme de visibilité, pour le producteur de l'information, est donc d'être présent dans ces vingt premières réponses. C'est un **enjeu stratégique et économique** majeur tout particulièrement pour les entreprises en terme de notoriété sur le marché et d'efficacité de leurs politiques marketing.

Nous verrons ultérieurement que ce positionnement est dépendant du mode de fonctionnement de l'outil de recherche et plus spécifiquement des

algorithmes de recherche utilisés, ce qui le rend très aléatoire et sans garantie de pérennité puisque les algorithmes des outils de recherche évoluent en permanence.

D'autre part des pratiques à caractère commercial se sont développées permettant aux producteurs d'information « d'acheter un positionnement » plus favorable. Cette évolution est la marque de l'enjeu économique que représente la présence et la visibilité sur le web et l'évolution du web vers la notion de coût de diffusion de l'information sur les réseaux.

Le *référencement* se définit comme la procédure permettant de proposer auprès des outils de recherche, moteurs et annuaires, un site (ou certaines pages d'un site) afin qu'il soit intégré dans leurs bases de données. La procédure de référencement sera différente en fonction du type d'outils visé (cf. infra).

Le référencement est donc la première étape, et une étape obligatoire, dans la stratégie de communication sur le web pour une entreprise, une administration, une association ou même un particulier. A ce niveau, l'enjeu est de donner aux internautes la possibilité **d'accéder à l'information**. En effet une information diffusée dans les pages d'un site web ne sera accessible que si celui ci a été repéré par les outils et intégré dans leurs bases de données fournissant les réponses aux requêtes des internautes. Si tel n'est pas le cas, ce site n'est accessible qu'aux personnes connaissant déjà son existence et son url (adresse web) ou éventuellement grâce à des liens pointant vers lui à partir d'autres sites. Ce qui limite fortement les chances d'être vu et les possibilités d'élargir son audience.

Le référencement auprès des outils de recherche est donc une première étape obligatoire à intégrer dans la démarche de promotion d'un site web. De plus, il est important d'intégrer cette démarche dès la phase de conception des pages d'un site. En effet, nous verrons dans le chapitre consacré au fonctionnement des outils de recherche, que l'indexation des pages web par un outil, ou la présence d'un site ou d'une page web spécifique dans les réponses fournies par l'outil à la requête de l'internaute, sont liés, entre autre, à différents éléments intégrés dans la page par le webmestre lors de la conception des

pages. Ces aspects sont souvent négligés, ce qui nuit à l'efficacité de la procédure de référencement de ces pages et donc à leur visibilité sur le web.

Le fonctionnement des outils de Recherche

Principes de fonctionnement des moteurs et incidences sur l'optimisation des pages :

La technique développée par les moteurs de recherche a éliminé l'intervention humaine dans le traitement de l'information pour n'utiliser que les capacités des ordinateurs à travailler, selon des technologies diverses (recherche en texte intégral, en langage naturel...), sur les textes de l'ensemble des documents mis en ligne sur le web. Le robot du moteur indexe le contenu d'une page web (texte de la page et balises méta associées), à partir de son code source, en extrait les mots clés (technique d'extraction), se constituant ainsi un catalogue ou index qui sera la base des réponses proposées à la requête de l'internaute. Le point fort de la recherche avec un moteur est de permettre une recherche extrêmement précise sur une chaîne de caractère. La recherche sera particulièrement pertinente sur des chaînes de caractères très discriminantes du type nom de famille ou terme scientifique ou technique.

Cependant la pertinence de la réponse est fortement liée à la qualité et à la précision de la requête de l'internaute. Hors ce niveau de qualité de la requête suppose que l'internaute sache utiliser les fonctionnalités de recherche avancée des moteurs mais également la syntaxe de saisie des requêtes propre à chacun de ces moteurs. C'est loin d'être le cas puisque la grande majorité des internautes effectuent leurs requêtes sur les moteurs en recherche simple avec un ou deux mots clés.

La conséquence directe d'une définition trop vague d'une requête est en retour plusieurs centaines voire milliers de réponses que l'internaute ne peut consulter. D'une manière générale, comme nous l'avons déjà dit, il se limite aux 20 à 30 premières réponses parmi lesquelles, il ne trouvera pas forcément l'information pertinente, même si celle-ci existe par

ailleurs sur le web mais exprimée avec un vocabulaire différent de celui de l'internaute.

La qualité de la requête conditionne donc fortement la pertinence de la réponse. C'est pourquoi l'optimisation des pages par le webmestre, et plus précisément des critères d'indexation, doit intégrer ces notions de comportement de l'utilisateur final et de vocabulaire adapté aux besoins de celui-ci, réalisant ainsi une réelle adéquation entre le contenu des pages d'un site et les attentes du public de ce site.

Les critères d'indexation

Les moteurs indexent généralement le texte des pages, les url (adresses des pages web, noms des fichiers) et les métadonnées incluses dans le code source. Les métadonnées sont des éléments apparaissant dans le code source d'une page web et fournissant des informations sur le contenu et la forme du document. Ces balises sont invisibles pour l'internaute et essentiellement destinées au traitement des pages par les outils de recherche. C'est pourquoi, elles sont un élément prépondérant dans la préparation du référencement d'un site sur les outils de recherche.

Il faut noter cependant que tous les moteurs n'ont pas le même comportement et que chacun d'entre eux développe son propre système de prise en compte et de valorisation des informations fournies par les métadonnées.

Les principales métadonnées prises en compte sont : **Titre, Description, Keywords, Attribut ALT des images** (légende ou commentaire apparaissant lorsque l'on passe sur l'image avec la souris), **robot txt**, (indication sur les conditions d'indexation par les moteurs, par exemple interdiction aux moteurs d'indexer certaines pages du site).

Chaque moteur développe donc sa technologie et la fait évoluer dans le temps. Ils prennent en compte très différemment les métadonnées (certains outils basent leur indexation sur l'ensemble des métadonnées, d'autres n'en utiliseront qu'une ou deux voire aucune). On peut remarquer cependant que la métadonnée « Titre de la page » semble

unanimement prise en compte par les moteurs, il est donc impératif de travailler le libellé des titres de chacune des pages d'un site afin d'augmenter la qualité de l'indexation de ces pages par les moteurs.

De nombreux sites consacrés au référencement annoncent que les métadonnées ne sont plus un élément déterminant de la visibilité d'un site sur le web, ce qui reste à prouver. De notre côté, nous pensons qu'il est essentiel, pour être bien référencé par les différents outils, d'optimiser **l'ensemble des métadonnées** dans le code source des pages les plus importantes du site au minimum, en prenant en compte les spécificités connues de chacun des outils pour obtenir un résultat maximum.

Le Dublin Core

En tant que professionnel de l'information nous recommandons d'ajouter à ces métadonnées une approche documentaire spécifique le « Dublin Core ».

Le « Dublin Core » (projet international initié en 1995) est un mode de marquage des pages avec des meta-tags spécifiques visant à homogénéiser les éléments essentiels des métadonnées afin d'optimiser la description des ressources accessibles sur le web. Les métadonnées développées par le « Dublin core » sont au nombre de 15. Elles font l'objet d'une norme ISO.

- Titre
- Auteur /créateur
- Sujet, mots-clés
- Description
- Editeur
- Contributeur
- Date
- Type de la ressource
- Format
- Identifiant de la ressource

- Source
- Langue
- Relation
- Couverture
- Gestion des droits

A l'heure actuelle, le « Dublin core » n'est que très peu pris en compte par les moteurs (Altavista, par exemple, semble prendre en compte, partiellement et à l'indexation, les métadonnées « Dublin core ») mais on peut penser que la situation peut être amenée à changer sur ce sujet et elle changera d'autant plus vite que le mouvement vers le « Dublin core » sera accompagné.

Parmi les quinze métadonnées développées par le « Dublin core » apportant un enrichissement des métadonnées existantes, on peut noter les plus couramment utilisées à ce jour : Auteur-créateur, éditeur, Date, type de ressource, gestion des droits.

Autres critères de pertinence

Au delà des ces critères d'indexation internes aux pages web (métadonnées et contenu de la page), les moteurs peuvent également utiliser d'autres critères du type "indice de densité d'un mot" (nombre de fois ou un mot est présent dans une page) ou encore "indice de popularité d'une page" (nombre de liens pointant vers cette page) ou encore "indice de clic" (nombre de fois ou les internautes cliquent sur un lien pointant vers une page parmi les pages proposées en réponse à une requête). Ces critères sont importants pour le positionnement dans la multitude des réponses fournies par un moteur à une requête. Par exemple, plus un mot-clé est présent sur une page (dans le contenu et dans les métadonnées) meilleur sera le positionnement de la page dans les résultats fournis par le moteur sur un requête posée sur ce mot-clé. Mais attention, la répétition excessive de mots-clés comme le font

certaines webmestres est considérée comme une tentative de tricherie et peut entraîner l'élimination définitive de l'index d'un moteur.

En revanche, et comme nous l'avons évoqué précédemment, les outils ont développé des propositions commerciales permettant par exemple d'acheter un positionnement sur un mot clé. Mais ici, nous sortons du champ du référencement et de l'indexation documentaires pour entrer dans le domaine du marketing.

En conclusion, il est donc important de prendre en compte l'ensemble de ces critères d'indexation afin d'optimiser des pages d'un site en vue de leur référencement, sans pour autant user de techniques abusives qui ne garantissent pas davantage un bon positionnement et une meilleure visibilité.

Principes de fonctionnement des annuaires

A l'inverse des moteurs, le répertoire est une **sélection de sites classés** par un opérateur humain au sein d'une classification pré-établie censée décrire l'ensemble des sujets traités sur le web.

Les sites sont répartis au sein d'une arborescence plus ou moins profonde en fonction du nombre de sites existants sur le web dans un domaine de connaissance ou d'activité déterminé, l'objectif étant que l'internaute accède, par navigation dans l'arborescence ou grâce au formulaire de recherche proposé, à une liste de sites réduite (quelques dizaines classés en général par ordre alphabétique) répondant à ses critères de recherche.

Pour les catégories les plus fournies en sites la classification prévoit généralement un classement géographique en plus du classement thématique.

Exemple : Dans le répertoire de Lycos dans la rubrique « informatique et recrutement » particulièrement fournie, on trouvera au 5^o niveau de l'arborescence un critère géographique « Paris »:

Economie, entreprise > Sociétés> Informatique, multi-média> ingénierie et conseil > Paris.

Pour chaque site présent dans la liste des réponses, les informations données sont le titre, l'url (adresse) et une brève description du contenu. Il est important de souligner que la réponse fournie à un internaute à partir des mots clés qu'il aura défini dans le formulaire de recherche, est une liste de sites web dont le contenu se rapporte au thème défini par les mots-clés de sa requête et non des pages web dans lesquelles on retrouverait ces mots précisément (type de résultats fournis par les moteurs de recherche).

Le référencement d'un site au sein d'un répertoire peut être le résultat d'une des trois démarches suivantes :

- le référencement d'un site ayant fait l'objet d'une demande de référencement payant (cas de plus en plus fréquent surtout dans le domaine économique et commercial),
- le référencement des sites soumis selon la procédure classique de référencement gratuit,
- la veille opérée par les personnes en charge d'un domaine du répertoire. Les documentalistes s'assurent ainsi par leur recherche qu'ils ne sont pas "passés à côté" d'un site important pour leur domaine.

Le référencement par soumission – gratuit ou payant – consiste donc à proposer son site dans la ou les catégories les plus pertinentes compte tenu de son contenu. Pour limiter le nombre de sites au sein des catégories, les répertoires imposent des règles (le site ne peut être soumis que dans une voire deux catégories, la catégorie retenue doit

être la plus précise possible, le site doit être régulièrement mis à jour et son contenu doit être suffisamment riche,...) et vérifient chaque demande de soumission, se réservant le droit de ne pas accepter le référencement ou de modifier la demande (changement de catégorie, modification du texte de présentation proposé par le site,...).

Le référencement d'un site dans un annuaire suppose donc pour un site de définir le plus précisément possible l'activité du site dans le texte de description en y intégrant les mots-clés susceptibles de correspondre au mieux aux termes employés par les internautes dans leurs recherches.

Pour des raisons économiques évidentes, les annuaires proposant des formules de référencement payant affirment ne plus consacrer que 30% du temps au référencement gratuit et les délais de prise en compte des demandes peuvent désormais être de plusieurs mois. Le principe de la procédure de référencement payant garantit une étude de votre site dans un délai rapide (8 jours environ) mais pas forcément son intégration dans la base de référence de l'annuaire.

La démarche d'OUROUK

Notre démarche, à l'inverse des sociétés qui se présentent comme spécialisées dans le référencement, est de type documentaire et axée sur un référencement professionnel de qualité et permettant de tendre vers un positionnement satisfaisant. En effet, le positionnement, en dehors des offres à caractère commercial, ne peut être garanti compte tenu des principes de fonctionnement évolutifs des outils de recherche évoqués précédemment.

Par ailleurs le référencement documentaire ne peut se concevoir que comme un **référencement manuel**. Les techniques de référencement automatique sur plusieurs dizaines d'outils, n'offrent aucune garantie de prise en compte réelle de votre demande sur les outils fondamentaux d'une part mais également sur les outils spécialisés dans un domaine précis susceptible de correspondre à la cible visée par le site.

Choix des outils : il faut savoir que l'essentiel du trafic est généré par une quinzaine d'outils parmi lesquels nous citerons les moteurs Google, Altavista, Voilà, MSN, Hotbot, Lycos, Fast... et pour les annuaires Yahoo, Nomade, Voilà (le Guide)... Il n'est pas forcément pertinent de consacrer du temps à référencer manuellement un site sur des outils peu utilisés par les internautes. Dans ce cas, et dans ce cas seulement, le recours à une procédure de soumission automatique auprès de ces outils secondaires peut être un complément intéressant. En revanche une soumission manuelle est essentielle en ce qui concerne les outils spécialisés qui connaissent actuellement un fort développement.

Enfin, les techniques de référencement automatique n'intègrent pas dans leurs procédures, les spécificités de fonctionnement et d'indexation de chacun des outils de recherche (critères et balises pris en compte) et de ce fait n'offrent pas les meilleures chances d'être référencé et bien positionné.

Par conséquent, notre démarche de référencement d'un site web propose :

1° le **référencement et son suivi** dans les outils de recherche. Nous n'offrons pas de démarche globale de publicisation d'un site qui allie l'action sur Internet et dans les autres médias (télévision, presse, etc.) ou uniquement dans l'Internet mais avec de multiples canaux.

2° l'emploi de **méthodes documentaires** pour référencer et ensuite retrouver un site. Les sociétés "spécialisées" dans le référencement pratiquent d'une manière ou d'une autre l'inflation de mots-clés (spamming) dans l'espoir de mieux positionner un site. Les méthodes sont diverses (pages satellites, ruses diverses pour tromper les moteurs, parmi les plus récentes on peut citer l'utilisation de la balise DIV, la répétition de mots-clés derrière certaines balises qui passent pour des paramètres non reconnus par le navigateur, etc.).

Ces démarches inflationnistes sont à courte vue. En effet les algorithmes des moteurs de recherche se modifient régulièrement et font de la chasse à l'inflation des mots-clés une de leurs préoccupations. Une des conséquences possible est la pénalisation des sites qui se livrent à de telles pratiques. Le résultat final est donc l'inverse de celui attendu car la pratique de l'inflation des mots clés ou l'introduction de mots-clés sans rapport avec le contenu des pages aboutit à mettre les sites qui les utilisent sur des listes noires et à rendre le référencement impossible.

3° La croissance de l'Internet est incomparablement plus rapide que le vocabulaire de la langue. Il en résulte que, mathématiquement, le nombre d'occurrence sur un mot clé croît. Cette croissance est d'autant plus importante que le mot-clé correspond à un mot générique employé régulièrement. Le **recours aux stratégies de recherche des professionnels de l'information** (association de plusieurs termes dans une recherche, recherche de mots-clés discriminants, recours à des recherches spécialisées) devient donc une nécessité toujours plus

grande. Dans cette perspective, l'indexation et le référencement des sites par des professionnels de l'information ont donc un intérêt croissant.

4° Ces évolutions conduisent les outils de recherche à effectuer des classements selon la notoriété des sites. La montée en puissance d'un site doit donc se placer dans le cadre d'une stratégie de long terme comme d'une stratégie de publicisation globale dans laquelle Ourouk peut se charger des aspects propres au référencement sur les outils de recherche.

5° La mondialisation de l'Internet s'accompagne de sa localisation. D'une part les divers univers linguistiques se développent, d'autre part, au delà de la langue, la logique nationale joue un rôle croissant qui conduit notamment les divers outils internationaux à avoir une offre adaptée à une nation tandis que se développent des concurrents locaux (par exemple, en France, le moteur de France Télécom, Voilà, est devenu incontournable). Le référencement doit **s'adapter aux divers univers, linguistiques, nationaux et thématiques des sites.**

Enfin si la visibilité sur le web est un enjeu primordial et le référencement un passage obligé, on ne doit attendre de ce dernier qu'une augmentation de trafic de 5 à 10%. Les internautes qui viennent par l'intermédiaire des outils de recherche sont aussi parmi les moins fidélisables.

IV Méthodologie du Référencement :

1° Audit du site :

- Architecture, contenu, code source, liens hypertexte
- Premières recommandations pour l'optimisation des pages.

2° Détermination d'une liste optimale d'outils de recherche

- Sélection d'outils généralistes
- Sélection d'outils à caractère géographique
- Sélection d'outils spécialisés

3° Analyse dans les annuaires de la présence du secteur thématique ou professionnel auquel appartient le site.

- Détermination des différentes catégories possible pour la soumission
- Analyse des descriptifs de présentation des concurrents ou autres organismes
- Proposition de choix de catégories de classement et rédaction du descriptif de présentation.

4° Analyse dans les moteurs de la visibilité des concurrents ou organismes équivalents

- Analyse des réponses fournies par les outils sur une liste de mots-clés spécifiques du domaine d'activité.
- Recommandations finales pour l'optimisation des pages web (balises métadonnées et organisation des contenus).

5° Soumission des URL aux moteurs et annuaires. Etablissement d'une fiche de référencement indiquant les moteurs et annuaires ciblés.

6° Suivi et relance du référencement sur les moteurs. Mise à jour de la fiche de référencement en fonction des résultats atteints. Dans cette prestation le

suivi de l'évolution de l'indexation du site une fois celui-ci accepté par le moteur ou le répertoire n'est pas assuré.

7° Suivi et relance, y compris par des moyens spécifiques (mél, etc.) jusqu'au moment où les outils de recherche ont pris en compte le site et au plus tard pendant un an.

Qui sommes-nous ?

Ourouk est une société de conseil dans le domaine des Nouvelles Technologies de l'Information et de la Communication créée en 1992.

Elle se développe aujourd'hui autour de 5 activités principales :

- ① Le conseil en stratégie, marketing et organisation des Nouvelles Technologies de l'Information et de la Communication :
- ② Le conseil en édition électronique et en conception de services interactifs :
- ③ Le conseil en management de l'information .
- ④ Des prestations associées à la gestion de votre information .
- ⑤ La formation .

Les consultants d'Ourouk disposent d'une expérience professionnelle de haut niveau dans les domaines des sciences de l'information, de l'Internet et des nouvelles technologies, de la documentation et du marketing. Ils sont issus du monde de l'entreprise et du conseil dans lequel ils ont exercé divers postes à responsabilité, notamment dans le domaine des NTIC. Ils maîtrisent efficacement les problématiques organisationnelles de l'entreprise.

La société est membre des associations professionnelles suivantes :

- . l'Association pour le Commerce et les Services en ligne (ACSEL),
- . l'Association des Professionnels de l'Information (ADBS),
- . l'Association des Professionnels de la Gestion Electronique des Documents (APROGED),
- . le Syndicat National des Prestataires et conseils en Information (SYNAPI).

L'EQUIPE

Ourouk se compose de 9 personnes, dont 4 collaborateurs pour la réalisation de prestations documentaires et 5 consultants seniors :

- **Pierre Brouste.** Docteur en Sciences Economiques. Expert auprès de l'AFNOR. Avant de créer Ourouk, Pierre Brouste a été administrateur de banques de données au Centre Français du Commerce Extérieur (CFCE) puis Directeur du Département Economie-Finances de SVP.
- **Myriel Brouland.** Diplômée de l'Institut d'Etudes Politiques de Paris et de l'Institut des Langues Orientales. Avant de rejoindre Ourouk, Myriel Brouland a été gestionnaire de banques de données à la Caisse des Dépôts et Consignations (CDC) puis Directrice des séminaires au Crédit Agricole et à la Fondation Royaumont.
- **Dominique Cotte.** Docteur en Sciences de l'Information. Maître de conférences en Sciences de l'Information et de la Communication à l'Université de Lille-III. Avant de créer Ourouk, Dominique Cotte a été responsable de la documentation de l'Express, puis Directeur de l'Information et de la Documentation du groupe BDDP.
- **Sylvie Cabral.** Diplômée de l'Ecole de Bibliothécaire Documentaliste et de l'ESCP (STEGE Direction Marketing). Avant de rejoindre Ourouk, Sylvie Cabral a été documentaliste à Centre INFFO, chargée d'études et chef de produit à la Direction Marketing du PMU puis co-fondatrice du cabinet Noème Consultants spécialisé en marketing des NTIC.
- **Laurent Bouvier-Ajam.** Avant de rejoindre Ourouk, Laurent Bouvier-Ajam a été Directeur associé de TRIEL, agence conseil en communication électronique, responsable du Département des études et du développement marketing du PMU et co-fondateur du cabinet Noème Consultants spécialisé en marketing des NTIC.

Au cours du IV^e millénaire avant Jésus-Christ, les entrepreneurs et les marchands de la ville d'Ourouk, en Mésopotamie, se posaient déjà le problème d'identifier des objets au moyen de signes durables et communiquer ainsi les données dont ils disposaient. Ils déposèrent des empreintes sur des plaques d'argile, donnant ainsi naissance à l'écriture.