

## Termes et définitions liés à l'internet, au web et aux NTIC. Les mots en **bleu** sont les équivalents anglophones.

**Annuaire (ou répertoire)/Directory** : Listes de sites web (et non de pages comme pour les moteurs), classés et hiérarchisés dans des catégories thématiques, organisée en arborescence. Ce travail de référencement est effectué par des humains qui jugent et classent ces sites périodiquement. Principaux annuaires : Yahoo!, ODP, Nomade...

**Browser / Navigateur** : Logiciel qui permet de surfer/naviguer sur le web et de l'explorer. Via une interface ergonomique, l'internaute peut parcourir les pages qui compose le web. Les principaux navigateurs du marché sont Microsoft Internet Explorer et Netscape Communicator et Navigator.

**Cookie / Témoin** : ce n'est pas un gâteau mais un fichier que le serveur d'un site web glisse (le plus souvent à votre insu), au sein de votre disque dur (Fichiers temporaires ou Dossier Cookies). Ce petit code stocke de l'information et mémorise vos visites. Il permet, par exemple de vous identifier lorsque vous revenez visiter un site régulièrement

**Cloacking** : technique qui consiste à soumettre aux spiders des moteurs, une page qui leur est spécialement destiné mas qui est différente de celle que verra l'internaute. Pour ce faire, on utilise un logiciel qui contrôle les adresses IP et qui reconnaît si-celle ci émanent d'un internaute humain ou d'un spider de moteur. Le plu souvent, les pages destinés aux moteurs utilisent les procédés de spamdexing pour améliorer le positionnement dans les pages de résultats.

**Clustering** : Sur un moteur ou métamoteur, le clustering consiste à classer et regrouper automatiquement les résultats dans des catégories homogènes dont les pages web vont avoir un ou plusieurs mots clés en commun. La clusterisation permet notamment de découvrir les différentes dimensions induites par une thématique ou un mot clé. Exemple : les métamoteurs Grokker, Clusty, Polymeta.

**CPC (Coût par Clic)** : système qui consiste à faire payer les annonceurs pour afficher des liens (appelées généralement liens sponsorisés) présentés sur les pages de résultats des outils de recherche, l'annonceur paye uniquement si l'internaute clique sur le ou les mots.. L'achat d'un mot clé par l'annonceur s'effectue selon un système d'enchères en ligne. Les sociétés de positionnement payant (comme Overture et Espotting) achètent aux moteurs de recherche l'exclusivité sur certains mots-clés. Ces mots sont ensuite vendus aux enchères aux annonceurs. Ainsi par exemple, Amazon achète à Yahoo! le terme "Harry Potter" pour obtenir un lien vers son site dans les premières réponse, à chaque fois que l'internaute tape ce terme sur Yahoo!

**CPM** : Coût d'une campagne publicitaire pour milles affichages d'une publicité mille. L'annonceur achète des liens (mots-clés), qu'il paye pour mille affichages, que les liens soient cliqués ou non.

**Forums de discussion/ Newsgroups** : service du net qui permet à un groupe de personnes d'échanger opinions ou idées autour d'une thématique, en direct ou en différé. Chaque internaute peut lire les messages des autres membres et apporter sa contribution. Les forums ont les plus connus sont ceux de Usenet. Les forums de discussion sont accessibles via un logiciel de messagerie (fonction "News") ou via le web.

**Index** : l'index d'un moteur ou annuaire de recherche représente la quantité de documents de l'internet enregistrée sur ses serveurs. Exemple : Google indexe plus de 4 milliards de documents net (pages web, images, archives des forums de discussion) sur plus de 8000 serveurs.

**Indexation/ referencig** : référencer (enregistrer) un document de l'internet (web, archives de newsgroups, images) dans un annuaire ou moteur de recherche.

**Liste de diffusion / Mailing list** : Discussion spécialisée sur un thème donné, diffusé par le biais de mails. Après inscription, chaque internaute peut participer à la discussion en envoyant un mail. La plupart des mailingls lists sont modérés c'est-à-dire que les envois sont contrôlés avant diffusion.

**Mailing list/ Liste de diffusion** : Discussion spécialisée sur un thème donné, diffusé par le biais de mails. Après inscription, chaque internaute peut participer à la discussion en envoyant un mail. La plupart des mailingls lists sont modérés c'est-à-dire que les envois sont contrôlés avant diffusion.

**Métamoteur / Meta search engine** : Outil de recherche qui, lors d'une requête, interroge simultanément plusieurs moteurs de recherche et annuaires/répertoires, classe les résultats par pertinence et en élimine les doublons.. Certains métamoteurs catégorise les résultats par technique en regroupant les réponses par dossiers thématiques ([Vivisimo](#)). Les métamoteurs sont disponibles en ligne sur le web ([Debriefing](#), [Ixquick](#), [Surfwax](#)...) ou sous forme de logiciels ([Copernic](#)).

**Moteur de recherche/ Search engine** : Système d'interrogation de bases de données. Un robot (spider) parcourt les documents présents sur le web pour les indexer (mettre en mémoire) sur des serveurs. Lors d'une recherche sur un

moteur, l'internaute lance une requête sur les bases de données de ces serveurs qui contiennent des millions de pages web. Principaux moteurs : Google, Altavista, Voila, AllTheWeb.

**Navigateur / Browser** : Logiciel qui permet de surfer/naviguer sur le web et de l'explorer. Via une interface ergonomique, l'internaute peut parcourir les pages qui compose le web. Les principaux navigateurs du marché sont Microsoft Internet Explorer et Netscape Communicator et Navigator.

**PAP** : Page avec publicité : bandeau, fenêtre pop-up, logo...

**P2P / Peer-to-peer** = poste à poste, "d'égal à égal". Avec ce type d'architecture de poste à poste, les ordinateurs (PC de simples particuliers. De 2 à plusieurs milliers) sont connectés les uns aux autres sans avoir besoin d'être reliés à un serveur central pour échanger des données. Les utilisateurs vont pouvoir échanger ainsi des fichiers mis à disponibilité sur les disques durs de chaque internaute du réseau. Le P2P permet de partager de nombreux types de fichiers (texte, vidéo, musique, image). Les applications P2P les plus connues sont les programmes de "file sharing" (partage de fichiers) musicaux qui utilisent différentes variantes de ces protocoles P2P. ex : [Kazaa](#), [eMule](#), Morpheus, Gnutella.

**Positionnement payant** : voir [CPC](#) et [CPM](#)

**Répertoire (ou annuaire)/directory** : Listes de sites web (et non de pages comme pour les moteurs), classés et hiérarchisés dans des catégories thématiques, organisée en arborescence. Ce travail de référencement est effectué par des humains qui jugent et classent ces sites périodiquement. Principaux annuaires : Yahoo!, ODP, Nomade...

**Robot / Spider** : module d'un moteur de recherche qui parcourt le web et l'internet (newsgroups, archives de mailings lists) afin d'en mémoriser les documents (URL, mots clés, corps du texte) pour alimenter l'index des moteurs.

**RSS (RDF Site Summary ou Rich Site Summary ou Real Simple Syndication)** : Un fichier RSS est un fichier [XML](#) qui permet de publier des titres de nouvelles ou d'articles formatés donc récupérables et exploitables dynamiquement par d'autres sites (syndication de contenu) via des [agrégateurs de contenu](#). Cete technologie est destinée à produire automatiquement des flux (fils) de "billets" ou d'actualités publiés avec leur texte, titre, date, l'auteur (sans mise en forme). On trouve ce format ur les blogs, les portails d'actualités (BBC, CNN, Yahoo!,...), les sites qui proposent des actualités régulièrement mises à jour (sites d'entreprises, de presse). Ces fils (feed) sont repérables via des icônes(clic-dessous) ou les termes : "syndiquez ce contenu", "syndication", "version XML"

**Spamdexing** : Contraction de spam (inonder une boîte mail de publicité) et index (l'index d'un moteur de recherche). La technique (répréhensible) du spamdexing a pour but de tromper les moteurs pour apparaître dans les premières places des résultats. Le webmaster va pour ce faire :-répéter de manière abusive des mots dans ses le corps de ses pages ou dans le code source (Balises TITLE, META description, META keywords)

- multiplier les noms de domaines pointant sur un seul et même site
- créer des pages uniquement pour augmenter la popularité en y multipliant les liens pointant vers le site.
- soumettre un site de manière répétitive à un moteur de recherche.
- faire du cloacking
- etc...

**Semantic Web / Web Sémantique** : Le [Semantic Web](#) est un concept développé en 1998 par Tim Berners-Lee. Dans "[Semantic Web Road Map](#)", il propose une nouvelle phase de développement du Web permettant d'améliorer l'exploitation des informations. D'après Tim Berners-Lee, le web "sémantique" est la possibilité d'obtenir une information organisée, normalisée de façon à rendre le web compréhensible aux machines afin d'en faciliter l'exploitation par les utilisateurs. Tim Berners-Lee est également le créateur du [W3C](#) (le World Wide Web Consortium) organisme officiellement chargé de la normalisation de tout ce qui concerne le web, dont le but est de promouvoir les actions et les techniques visant à l'avènement du Web Sémantique. Le W3C a ainsi développé plusieurs langages pour manipuler les contenus. En effet, RDF (Resource Description Framework) et XML (eXtensible Markup Language), définis par le W3C constituent les bases du Semantic Web. [Web sémantique, philosophie et standards](#).

**Spider / Robot** : module d'un moteur de recherche / qui parcourt le web et l'internet (newsgroups, archives mailings lists) afin d'en mémoriser les documents (URL, mots clés, corps du texte) pour alimenter l'index des moteurs. On parle aussi de Crawler (ramper)

**Streaming** : Technologies de diffusion du son ou de vidéo en flux continu sur le Web. Permet de lire les fichiers au fur et à mesure du transfert sans avoir à attendre la fin de leur téléchargement

**Tag** : Etiquettes. Choies et attribuées librement par les internautes, ce sont des mots clés qui permettent de définir, classer puis rechercher des documents, des favoris, des images. Approche subjective sans organisation, hiérarchie et terminologies préétablis comme sur les systèmes documentaires. Exemple : les tags définissant les favoris sur [del.icio.us](http://del.icio.us) (partage de favoris)

**Nuage de Tags** : affichage d'un groupe de tags permettant de visualiser les plus fréquemment employés. La taille des mots est proportionnelle à leur fréquence d'utilisation. Plus un mot est gros, plus il est employé. Les tags sont employés dans le cadre de la [Folksonomie](#).

**Web** : World Wild Web (Toile d'araignée) : composante de l'internet. Structure de navigation sur un réseau de pages via des liens hypertextes, mots sur lesquels il suffit de cliquer pour accéder à un autre document.

**Web 2.0** : nouvelle génération des applications web. Une génération qui offre des outils pour partager de l'information sur le net d'une façon à la fois communautaire et personnalisée. Là où l'internaute était passif et statique avec le Web 1.0, il est devenu actif, participatif et socialisant avec ce Web 2.0. L'internaute crée du contenu, le partage, l'organise, le personnalise.

- Applications et principes types : bookmarks sociaux, pages d'accueil personnalisables, partage d'images, moteur de recherche à périmètre personnalisable, réseaux sociaux, blogs, RSS, mashups, bureautique en ligne, moteurs de tags, [folksonomie](#),....

Les [tags](#) permettent d'étiqueter et rechercher le contenu.

- Technologies associées : XML, Ajax, DHTML, Soap...

**Web invisible / Invisible web, deep web, hidden web** : Le "web invisible" désigne la partie du web non accessible aux moteurs de recherche classiques.

Et ce parce que:

- les documents sont trop volumineux pour être entièrement indexés,
- des pages sont protégées par l'auteur (balise meta qui stoppe le robot),
- des pages sont générées seulement dynamiquement, lors d'une requête par exemple,
- le format des documents n'est pas reconnu par les moteurs (de moins en moins vrai avec un moteur comme Google...).

**Blog, Blogue, Carnet**: Page perso mis à jour régulièrement sous forme d'actualités datées, signées avec possibilité d'ajouter un commentaire pour les visiteurs. C'est une sorte de journal de bord qui va de la confession intime à une analyse de l'actualité. Les articles sont présentés par ordre déchronologique

**Wiki/ Wiki**: Les Wikis sont des sites web dynamiques dont tout visiteur peut éditer et modifier de façon simple et rapide n'importe laquelle de ses pages. Créé en 1995 par Ward Cunningham, le Wiki tire son nom du mot hawaïen "WikiWiki", qui signifie "vite". Aujourd'hui utilisé par des communautés professionnelles ou non, pour collaborer sur des projets collectifs

**XML (eXtensible Markup Language)** : langage qui permet à un créateur de pages HTML de définir ses propres marqueurs (tags) et donc de personnaliser la structure et l'organisation des données qu'il présente. Ex :

```
h1><prenom>Steven</prenom>
<font color="red"><nom>Soderbergh</nom></font></h1>
<naissance>1963</naissance>
```

Cela permet une meilleure définition, transmission, validation et interprétation des données entre applications et organisations. Les [weblogs](#) sont développés en XML.