



GUIDE PRATIQUE DE LA RECHERCHE D'INFORMATIONS SUR INTERNET

JUIN 2009

CHAMBRE DE COMMERCE ET D'INDUSTRIE DE LYON

- Sommaire –

Améliorez votre recherche d'informations sur Internet.....	3
1. Typologie des outils de recherche sur Internet	4
2. Les outils de recherche du Web visible	5
2.1. Les moteurs de recherche.....	5
2.2. Les métamoteurs.....	6
2.3. Les annuaires.....	6
2.4. Les fils RSS.....	7
3. Les outils de recherche du Web invisible.....	8
3.1. Les moteurs de recherche du web invisible	8
3.2. Les répertoires sélectifs.....	9
3.3. Les outils de recherche thématiques.....	9
3.3.1. Les annuaires thématiques ou portails sectoriels	9
3.3.2. Les sites fédérateurs	10
4. Syntaxe d'interrogation.....	11
4.1. Principe de formulation.....	11
4.2. Fonctions avancées des moteurs de recherche les plus utilisés	11
5. Pour aller plus loin... ..	12

Améliorez votre recherche d'informations sur Internet

Internet véhicule à l'échelle mondiale une immense quantité d'informations et représente une mine de renseignements pas toujours faciles à identifier. Pour mener une recherche efficace, il est nécessaire de consulter une méthodologie et de connaître les différents outils.

Il existe de nombreux outils accessibles en ligne permettant de sonder le Web et d'identifier l'information dont vous avez besoin. Ces outils n'utilisent pas tous le même fonctionnement, aussi il est utile d'en connaître les principes.

Il y a une distinction fondamentale à faire entre les outils de recherche disponibles sur Internet : les outils du Web visible et les outils du Web invisible. L'information contenue sur Internet n'est pas accessible de façon homogène, il existe différents langages, différentes méthodes pour la collecter.

Le Web visible regroupe les pages web indexées par les moteurs, métamoteurs et répertoires, ce sont des pages dites « statiques », qui ne se modifient pas lorsqu'un internaute les consulte. Ces pages sont accessibles facilement grâce aux outils classiques de recherche.

Le Web invisible représente 80% de la masse d'informations contenue sur Internet et n'est donc pas à écarter dans le cadre d'une recherche d'informations. Il regroupe les pages Web dites « dynamiques », les outils classiques ne peuvent pas les répertorier. Ces pages sont soit non localisables (la page Web ne se constitue que lorsque l'internaute l'interroge), non référencées ou construites à partir d'un langage spécifique que les outils de recherche classiques ne reconnaissent pas.

"Les ressources du Web invisible comprennent, entre autres, les sites Web construits autour d'une base de données (interrogeable uniquement par un moteur de recherche interne), les pages accessibles par un formulaire de recherche, les pages protégées par un mot de passe, les pages interdites aux robots d'indexation, [...]."

L'espace Intelligence Economique vous propose ici une typologie des outils de recherche ainsi qu'une méthode pour élaborer une bonne stratégie de recherche sur Internet (comment rechercher de l'information sur un secteur d'activité, un produit spécifique, une entreprise, etc.).

1. Typologie des outils de recherche sur Internet

	Description	Spécificité	Mode de recherche	Type d'information
Outils de recherche du web visible				
Moteur	Base de données de pages Web, classées dans un index Réalisé par un robot	Liste de pages Web, texte intégral Répond à des questions précises Classement des résultats par ordre de pertinence	Par mot-clé	Généraliste Spécialisée
Métamoteur	Outil de recherche sur plusieurs moteurs simultanément	Recherche simple et précise sur des mots clés spécifiques	Par mot-clé	Généraliste Spécialisée
Annuaire	Catalogue de sites concernant un thème donné Réalisé par une personne	Liste de sites Internet Identification de sources spécialisées Liste de sites similaires par exploration de la rubrique	Par thème Par mot-clé	Généraliste Spécialisée
Outils de recherche du web invisible				
Répertoire sélectif	Répertoire de sites classés par thèmes, réalisé par une personne	Liste de sites Internet (quelques milliers maximum), largement ouvert sur le Web invisible	Par thème Par mot-clé	Spécialisée
Annuaire thématique Portail sectoriel	Répertoire de sites concernant un thème, un secteur d'activité, une filière	Deux types de portail : - Horizontal (droit, marketing, etc.) - Vertical (secteur d'activité)	Par thème Par mot-clé	Spécialisée
Site fédérateur	Portail thématique avec un aspect marchand	Annuaire spécialisé, informations pratiques, échanges et espace de fidélisation	Par thème Par mot-clé	Spécialisée
Page de liens	Liste de liens hypertextes sur un sujet donné dans des sites Internet	Accessible via les moteurs de recherche du Web invisible avec une syntaxe particulière	/	Spécialisée
Autres outils				
Blog	Page Web publiée par un internaute	Information publiée par des internautes (amateurs ou professionnels), interactivité, Utile en veille concurrentielle pour détecter les rumeurs	Par mot-clé	Généraliste Spécialisée
Fil RSS	Fichier texte dont le contenu est produit automatiquement en fonction des mises à jour d'un site Web.	Utilisé par les organes de presse. Permet de suivre l'actualité en temps réel sans avoir besoin de se connecter au site concerné.	Par mot-clé	Généraliste Spécialisée
Forum de discussion	Echange et consultation de messages en différé, nécessite généralement une inscription (gratuite ou payante)	Utile en veille concurrentielle pour détecter les rumeurs Les forums comprenant un modérateur (qui filtre les messages) sont un gage de qualité	Par mot-clé Par arborescence de sujets	Spécialisée
Liste de diffusion	Echanges par courrier électronique, nécessite généralement une inscription (gratuite ou payante)	Suivi de l'actualité d'un secteur (veille) Les listes comprenant un modérateur (qui filtre les messages) sont un gage de qualité	Par mot-clé quand les archives sont disponibles	Spécialisée
Wiki	Site Web collaboratif où chaque internaute visiteur peut participer à la rédaction du contenu.	Mutualisation des connaissances	Par mot-clé Par rubrique	Généraliste Spécialisée

2. Les outils de recherche du Web visible

Le Web visible peut être interrogé à partir de trois types d'outils : les moteurs de recherche, les métamoteurs et dans une moindre mesure les annuaires.

2.1. Les moteurs de recherche

Le moteur de recherche est un outil entièrement automatisé qui utilise un robot pour collecter le contenu des pages Web et les stocker dans une base de données. Cet outil permet donc de rechercher en texte intégral dans les pages Web de plusieurs sites.

<i>URL</i>	<i>Couverture</i>	<i>Taille</i>	<i>Fréquence d'actualisation</i>
Moteurs d'origine anglophone			
Google en anglais Google en français	Mondiale	60 milliards de pages	Continuel
Yahoo ! Search en anglais Yahoo ! Search en français	Mondiale	6,6 milliards de pages	Continuel
Bing en anglais Bing en français	Mondiale	5 milliards de pages	Continuel
Ask en anglais Ask en français	Mondiale	5,3 milliards de pages	4 semaines (variable)
Cuil	Mondiale	124 milliards	continuel
Gigablast	Mondiale	4 milliards	continuel
Moteurs d'origine francophone			
Exalead	Européenne	8 milliards de pages	4 semaines (variable)
Voilà	Francophone	60 millions de pages	4 semaines (variable)

Pour interroger les moteurs de manière pertinente, il est utile de connaître leur syntaxe d'interrogation. Pour cela, vous pouvez notamment consulter le site [Abondance](#) : une notice détaillée est disponible pour chaque outil.

Par ailleurs, un outil accessible en ligne permet de comparer la couverture des moteurs de recherche : [Thumbshots Ranking](#).

2.2. Les métamoteurs

Le métamoteur est un outil qui interroge simultanément plusieurs moteurs de recherche, annuaires ou banques de données. Il propose en résultat les 10 premiers sites ou pages trouvés pour chaque outil interrogé.

<i>URL</i>	<i>Couverture</i>	<i>Outils interrogés</i>	<i>Indication de la source</i>
Métamoteurs généralistes anglophones			
Clusty	Mondiale	18 outils	Oui
Ixquick	Mondiale	12 outils	Oui
Dogpile (leader mondial)	Mondiale	4 outils	Oui
AllPlus	Mondiale	4 outils	Oui
Métamoteurs généralistes francophones			
Kartoo	Mondiale	14 outils	Non
Ariane6	Francophone	14 outils	Oui

Il existe également des métamoteurs spécialisés dans un secteur, une thématique. Ils sont pour le moment peu nombreux et assez mal répertoriés. Les noter au fur et à mesure de vos recherches est une bonne technique pour les lister.

2.3. Les annuaires

Un annuaire est un répertoire de sites Internet, évalués, décrits et classés par une personne dans des rubriques. L'annuaire ne recense que l'URL du site et un descriptif de son contenu. La recherche porte donc uniquement sur les éléments recensés.

<i>URL</i>	<i>Couverture</i>	<i>Taille</i>
Annuaire d'origine anglophone		
Yahoo ! en anglais Yahoo ! en français	Mondiale	Environ 2 millions de sites en anglais, 150 000 sites en français
Dmoz en anglais Dmoz en français	Mondiale	Environ 4 millions de sites en anglais, 220 000 sites en français
The Internet Public Library	Mondiale	Environ 40 000 sites en anglais
Annuaire d'origine francophone		
NEToo	Francophone	Environ 10 000 sites en français
Indexa	Francophone	Environ 60 000 sites en français

2.4. Les fils RSS

Un fil RSS est un fichier texte, généré automatiquement, contenant la description des nouveautés mises en ligne sur un site web, ainsi que le lien pointant vers le contenu de ces nouvelles informations. Ainsi, il permet de suivre l'actualité en temps réel sans avoir besoin de se connecter au site concerné. Lorsqu'un site dispose d'un flux RSS, on le reconnaît par un des symboles suivants :



Pour pouvoir lire et s'abonner à des flux RSS, il faut un lecteur ou agrégateur de flux RSS.

Il existe trois formes de lecteurs de flux RSS :

- les lecteurs intégrés au navigateur Internet : Mozilla Firefox, Internet Explorer (à partir de la version 7.0).
- les lecteurs « en ligne » : [Netvibes](#), [Bloglines](#) (en anglais), [Google Reader](#).
- les lecteurs sous forme de logiciel : Alertinfo, RSS Bandit, FeedReader, Feeddemon.

Plusieurs logiciels de messagerie comme Outlook 2007 ou Mozilla Thunderbird, permettent également de collecter les flux RSS.

3. Les outils de recherche du web invisible

Le Web invisible est accessible grâce aux outils suivants : les moteurs de recherche spécialisés, les répertoires sélectifs, les outils de recherche thématiques (annuaires thématiques, portails sectoriels, sites fédérateurs, pages de liens, etc.).

Les outils de recherche thématiques et sélectifs sont nombreux et nous ne citerons que quelques sources à titre d'exemple, d'autres étant présentés dans l'annuaire de sites « 1000 sites ».

3.1. Les moteurs de recherche du web invisible

Afin d'identifier les sites du Web invisible, il convient d'utiliser des moteurs, métamoteurs ou répertoires non conventionnels. Voici quelques outils qui pourront vous aider à débiter vos recherches.

<i>URL</i>	<i>Couverture</i>	<i>Type d'informations</i>
Moteurs de recherche		
Yahoo ! Search	Mondiale	Publications de référence, d'études et de grands serveurs commerciaux (payant)
Google Scholar	Mondiale	Documents académiques, livres blancs et articles scientifiques
Métamoteur		
Scirus	Mondiale	Mémoires, thèses, rapports et papiers de recherche
Répertoire		
Find Articles	Mondiale	5 millions d'articles

3.2. Les répertoires sélectifs

Les répertoires sélectifs sont des outils créés et alimentés par des personnes, bien souvent professionnels de l'information. Ce sont des listes de sources jugées pertinentes dans leur domaine et sont généralement ouvertes sur le Web invisible. L'internaute pourra y trouver des liens vers des bases de données, des sites de références, la presse, les listes de diffusions, les documents de recherche.

<i>URL</i>	<i>Ressources</i>
Répertoire de signets	
Bibliothèque Nationale de France	3400 ressources - 800 rubriques
Bibliothèque Georges Pompidou	47 rubriques
Ecole Sciences Politiques	1000 liens - 40 bases de données
Annuaire « 1000 sites »	Rubrique <i>Répertoires d'entreprises et annuaires</i>
Rime	Ressources en management et économie 650 sites - 18 rubriques
Infomine	9 rubriques (dont Sciences, Economie, Business, Santé, etc.) En anglais
Intute	4 rubriques (Sciences et technologie, Art, Sciences sociales, Santé et médecine) En anglais
Répertoire de bases de données	
Dadi	1120 bases de données gratuites
Complete Planet	70 000 banques de données

3.3. Les outils de recherche thématiques

3.3.1. Les annuaires thématiques ou portails sectoriels

Pour trouver une information spécifique à une filière ou à un secteur d'activité, il existe des annuaires thématiques ou portails sectoriels. Pour trouver ce type d'outils, vous pouvez consulter des « portails de portails », c'est-à-dire des répertoires ou annuaires de portails sectoriels. Vous trouverez une liste non exhaustive de ces outils dans la rubrique *Répertoires de sites* de l'annuaire « 1000 sites ». Vous pouvez également consulter ces deux outils : [Indexa](#) et [Les Annuaires](#).

3.3.2. Les sites fédérateurs

Les sites fédérateurs sont des outils proches des portails thématiques, mais avec un aspect marchand. Ils associent les acteurs et centralisent les informations et transactions d'un secteur particulier. Les sites fédérateurs sont généralement composés de quatre sections : un annuaire spécialisé, un contenu éditorial (information, interviews, dossiers), des échanges (forum, liste de diffusion, petites annonces) et un espace de fidélisation (espace membre, boutique, catalogue).

Ces outils sont localisables sur Internet grâce aux annuaires thématiques, répertoires de signets, etc. Vous pouvez également interroger le site [Search Engine Colossus](#), qui répertorie des moteurs et des annuaires par pays et par thèmes.

4. Syntaxe d'interrogation

Une recherche efficace sur un moteur de recherche suppose l'utilisation de certaines fonctions et formulations. Voici les éléments de syntaxe les plus utiles pour débiter une recherche.

4.1. Principe de formulation

- **Les opérateurs booléens :**

ET / AND / + : la recherche se fait sur deux mots et plus

OU / OR : la recherche prend en compte au moins l'un des mots

SAUF / NOT / - : La recherche exclut les pages contenant ce(s) mot(s)

- **Les guillemets (" ")** : pour retrouver une expression exacte

- **La troncature (*)** : tous les moteurs ne proposent pas cette fonctionnalité, il vaut mieux utiliser l'opérateur OR pour demander les singuliers et les pluriels.

- **Caractères accentués** : tous les moteurs ne les gèrent pas de la même façon. A priori, il vaut mieux saisir le terme avec accent et pour une recherche exhaustive comparer les résultats avec ou sans accent.

4.2. Fonctions avancées des moteurs de recherche les plus utilisés

	<i>Google</i>	<i>Yahoo ! Search</i>
<i>Opérateurs booléens</i>	<i>AND (par défaut)</i> <i>OR</i> <i>- (sauf)</i>	<i>AND (par défaut)</i> <i>OR (parenthèses)</i> <i>- (sauf)</i>
<i>Recherche sur champs</i>	<i>intitle:mot</i> <i>allintitle:plusieurs mots</i> <i>inurl:mot</i> <i>allinurl:plusieurs mots</i> <i>site:(gouv.fr)</i>	<i>intitle:mot</i> <i>TITLEWORDS(mot1 mot2)</i> <i>inurl:mot</i> <i>URLWORDS(mot1 mot2)</i> <i>site:</i>

A noter que les moteurs proposent des écrans de recherche avancée qui facilitent l'utilisation de ces champs. Par ailleurs, la syntaxe détaillée des différents moteurs de recherche est disponible sur le site [Abondance](#).

5. Pour aller plus loin...

Une rubrique de l'annuaire « 1000 sites » (*Management de l'information*) est entièrement consacrée aux ressources et producteurs d'informations. Vous pourrez y trouver des liens vers des bases de données (Dialog, Electre), des informations pratiques sur l'intelligence économique (les Arist, Veille Magazine), des revues spécialisées (Archimag), etc.

Il est important de se tenir informé de l'actualité des sources d'information et de l'évolution de ces outils. Les revues spécialisées [Bases et Netsources](#) (éditeur FLA Consultants) diffusent de nombreuses informations et conseils sur la recherche sur Internet (les archives de plus d'un an sont consultables gratuitement).

L'[Espace Numérique Entreprises](#), centre de ressources et de compétences sur les Technologies de l'Information et de la Communication au service des entreprises, publie des fiches pratiques très utiles sur les thèmes du Web, des nouvelles Technologies ou encore de la sécurité informatique.

CHAMBRE DE COMMERCE ET D'INDUSTRIE DE LYON Direction des Services aux Entreprises Espace Intelligence Economique Place de la Bourse 69289 LYON CEDEX 02	Tél. 0821 231 251 (0,12 € TTC/mn) http://www.lyon.cci.fr
--	--