

Internet utile pour le bibliothécaire

Support de formation

*juin 2009
Gap*

Sommaire

Les articles et fiches de ce support sont extraits des documents téléchargeables sur internet listés ci-dessous dans leur ordre de présentation

Extraits du diaporama utilisé durant la formation

- [<http://www.netvibes.com/savoirenactes>]

Fiches de présentation de la syntaxe d'interrogation des outils de recherche :

- **Google** : [<http://outils.abondance.com/google.html>]
- **Exalead** : [<http://outils.abondance.com/exalead.html>]

GUERTIN Hélène (dir.), *Chercher pour trouver*, EBSI, Univ. de Montréal

- [<http://www.ebsi.umontreal.ca/jetrouve/>]
- Les 6 étapes d'un projet de recherche d'information
- J'évalue un site web
- J'évalue le contenu d'une page web

NOEL Elisabeth (dir.), *Ressources Electroniques Pour les Etudiants, la Recherche et l'Enseignement*, ENSSIB, 2008.

[<http://repere.enssib.fr>]

- Techniques d'accès aux ressources électroniques et à l'Internet
- Gestion des signets, favoris ou bookmarks
- Les outils de recherche d'informations sur le web



CC By – NC Emily Barney (Flickr)

Internet utile pour le bibliothécaire

8 & 9 juin 2009 – Gap
Bibliothèque Départementale des Hautes-Alpes

SAPIENS 2009
21/03/09

Philippe CAZENEUVE (Oxalis / Savoir en actes)
www.savoirenactes.fr

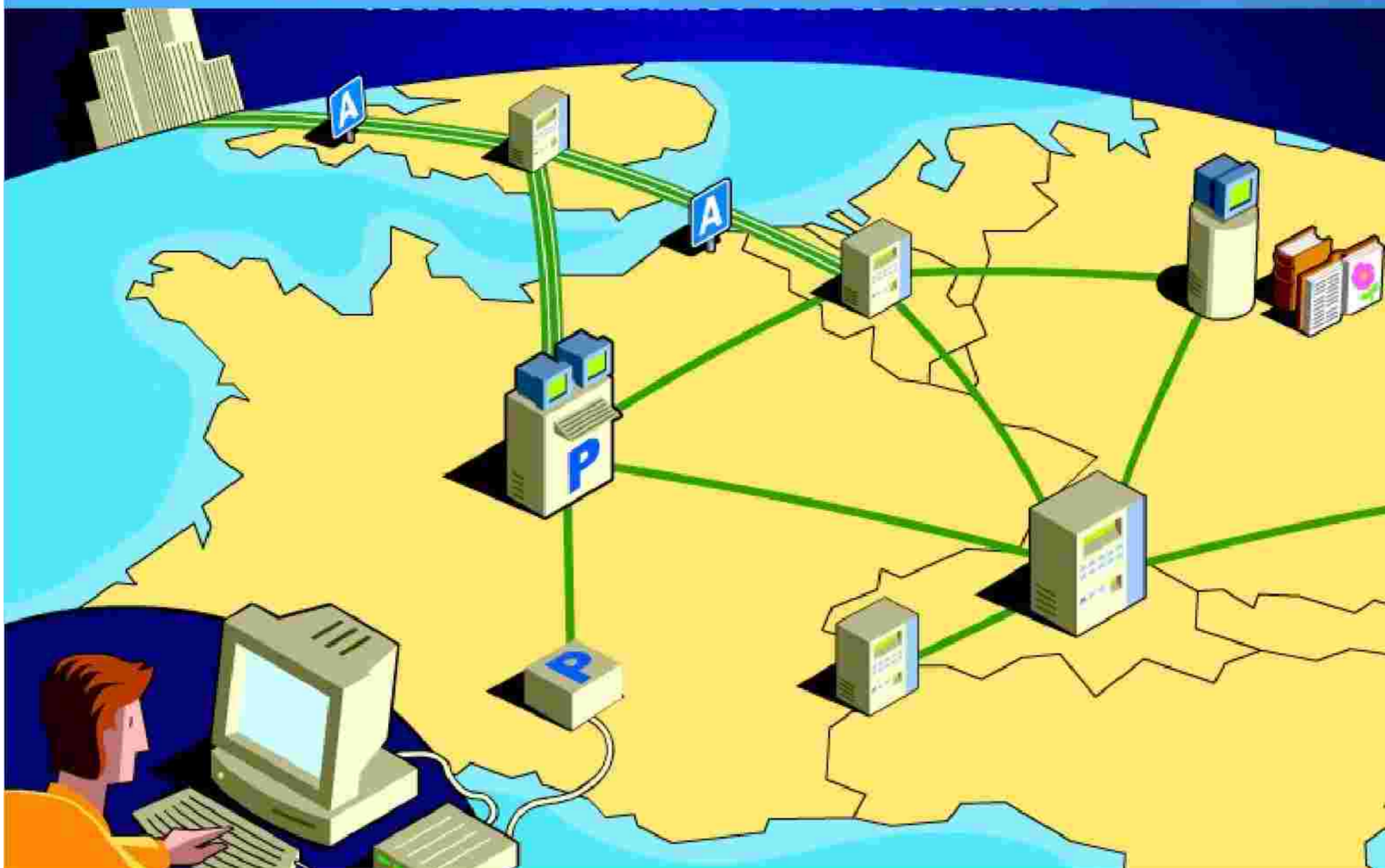
Retrouvez toutes les ressources de cette formation sur

www.netvibes.com/savoirnactes

The screenshot shows a Netvibes dashboard with the following widgets:

- Accueil**: Chercher & trouver sur Internet, Accueil des personnes handicapées, Biblio 24
- 1 - Introduction**: Un peu d'histoire, Le fonctionnement technique, Le modèle économique du développement de l'internet, Les évolutions en cours: web 2.0, bibliothèques numériques
- 2 - Classement des moteurs de recherche**: TOP 3 des moteurs de recherche en parts de marché (décembre 2008). Pie chart showing Google (64.2%), Yahoo! (12.2%), and Bing (1.2%).
- 3 - Potentiels des outils de recherche**: Annuaire thématiques et guides spécialisés, Moteurs de recherche, Méta-moteurs de recherche, Encyclopédies, dictionnaires, sites de référence, Recherche d'images, photos, Recherche de musique et vidéos, Partage collaboratif de signets (social bookmarking), Transports: Plans, Itinéraires, horaires, ...
- 4 - Signets de la BNF**: <http://signets.bnf.fr/> Les Signets de la Bibliothèque nationale de France
- 5 - Histoire de Recherche**: Search engines: Google.fr, Yahoo!, Excite
- 6 - 40 outils de recherche en 10 clics**: Mot(s) puis clic sur l'outil, SITES WEB, Google, Yahoo!
- 7 - Au cœur d'un moteur de recherche**: L'indexation automatique des pages: référencement, méta-données, robots d'indexation, Le tri et le classement des résultats affichés: Pagerank, popularité, liens sponsorisés..., Le web invisible: ce que les moteurs de recherche répertorient mal.
- 8 - Pour aller plus loin**: Des ressources pour approfondir
- 9 - RÈPERE - Ressources électroniques pour les étudiants, le**
- 10 - Tout ce que vous voulez savoir sur Google**: Tout ce que vous voulez savoir sur Google, Tout ce que vous avez toujours voulu savoir

Internet, Comment ça marche ?



Annuaire

Moteurs de recherche



Votre départ sur le Web

<http://www.fouineux.com>



<http://signets.bnf.fr/>



<http://www.dmoz.org/>



<http://www.google.fr/>



<http://www.exalead.fr/>

Annuaire

- Indexation par des humains
- Recherche sur des sites par catégories
- ++ Classement raisonné et sélectif
- -- Actualisation plus lente
- => Pour trouver des sites de référence sur un thème

Moteurs de recherche

- Indexation par des robots
- Recherche en texte intégral sur des pages web
- ++ Meilleure exhaustivité et actualisation plus rapide
- -- Information pertinente noyée dans la masse
- => Pour des recherches par mots-clés

Métamoteurs de recherche

Une seule requête simultanée sur plusieurs moteurs ou annuaires

- Application hors-ligne



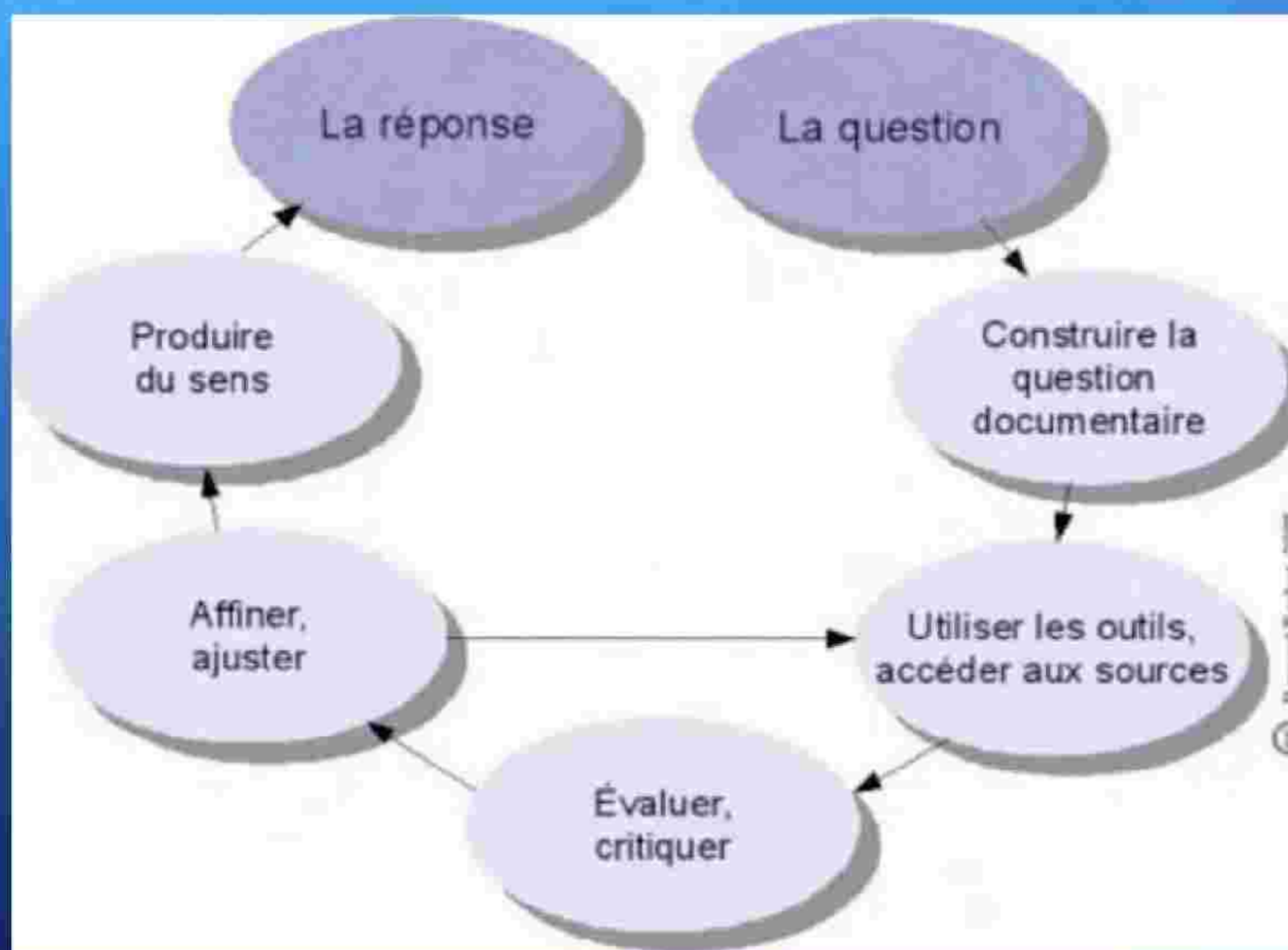
<http://www.copernic.com/>

- Site en ligne



<http://www.kartoo.fr/>

La boucle de recherche d'information





Toute l'info et l'actu sur les annuaires et moteurs de recherche : Recherche d'information et référencement

- [Vous débutez ?](#)
- [Actualité](#)
- [Blog](#)
- [L'actu sur votre site](#)
- [Dossiers / Articles](#)
- [Méthodologies](#)
- [Outils de recherche](#)
- [Audits](#)
- [Forums / Chat](#)
- [Lettres d'information](#)
- [Études](#)
- [FAQ](#)
- [Livres](#)
- [Emploi](#)
- [Tribune](#)
- [Humour](#)
- [Liens](#)
- [Offre commerciale](#)
- [Boutique en ligne](#)
- [Zone Abonnés](#)
- [Réseau Abondance](#)

[Accueil](#) > [Outils de recherche](#) >

[A la une de l'actu](#)

► Webmasters, **installez les titres de l'actualité** sur votre site !
► **Toute l'actualité de la semaine**

Google



Présentation générale

Nom	Google
URL	http://www.google.com/ http://www.google.fr/
Date de lancement	1998
Propriétaire	Google
Taille de l'index	8 milliards de pages
Délai de rafraîchissement de l'index	Continuel (dépend des délais de mise à jour des pages et de leur PageRank)
Nom du spider	Googlebot
Technologie utilisée par les portails :	Free, Club-Internet (France), Numericable, AOL.com, Earthlink, Netscape, etc.

Syntaxe d'interrogation

Minuscules / majuscules	Idem : Google ne tient pas compte de la casse des lettres. Exemple : ibm trouve ibm , IBM ou Ibm . Ibm trouve Ibm , ibm ou IBM . IBM trouve IBM , ibm ou Ibm .
Lettres accentuées	Important. electricite et électricité ne donnent pas le même résultat, même si les différences sont souvent minimes.
Ordre des mots	Important : paris dakar donne un résultat différent de dakar paris . Une plus grande importance est donnée au premier mot choisi.
OU	OR Exemple : engine OR directory . L'opérateur doit être saisi en majuscules obligatoirement.
ET	Opérateur par défaut Exemple : moteur recherche recherche les pages qui contiennent à la fois moteur ET recherche . Il est également possible d'utiliser le signe + pour demande rune orthographe spécifique : Exemple : +jérôme ne trouvera pas la forme "jerome" (non accentuée)
SAUF	- Exemple : moteur -automobile recherche les pages qui contiennent moteur mais qui ne contiennent pas automobile .
Expressions	Guillemets. Exemple : "moteur de recherche" recherche cette expression, les trois mots les uns à côté des autres et dans cet ordre.
Troncature	Non Il n'est pas possible de faire des recherches en utilisant la troncature sur Google. Le moteur recherche toujours exactement le mot demandé. mot ne trouve pas mots ni moteur . L'astérisque (*) ne peut pas être utilisé. Google tient cependant parfois compte de la troncature, sans qu'il soit possible pour l'internaute de décider quand.
Synonymes	Tilde (~) Une recherche sur powerpoint ~help effectuera une recherche sur powerpoint ET help ou tips , faq , tutorial , etc.
Recherche sur le nom de domaine du site	site: Exemple : +moteur +site:www.honda.com recherchera les documents qui contiennent le mot moteur et qui sont disponibles sur le site www.honda.com .
Recherche sur	link: Exemple : link:www.abondance.com trouvera les documents qui contiennent

Chercher sur le Web

avec  mozbot :

Recherche

Chercher sur :

- Web international
- Web francophone

REFERENCEMENT 2.0

Olivier Andrieu

Le livre sur le référencement de sites web



les adresses des liens	un lien vers une page du site dont l'adresse contient l'expression www.abondance.com (dans ce cas www.abondance.com).
Recherche de sites similaires	related: Exemple : related:www.abondance.com trouvera des sites similaires à Abondance.
Recherche dans le cache	cache: Exemple : cache:www.abondance.com trouvera la version du site Abondance dans le cache du moteur.
Informations sur le site	info: Exemple : info:www.abondance.com proposera un certain nombre d'informations (pages similaires, pages du site, pages similaires...) sur le site www.abondance.com .
Recherche sur la date	daterange: Exemple : google daterange:2452640-2452779 . Les dates doivent être indiquées au format Julien.
Recherche sur le type de fichier	filetype: Exemple : css filetype:pdf . Plusieurs formats sont possibles.
Recherche sur les mots	define: Exemple : define:engine trouvera sur le Web des définitions du mot engine .
Recherche d'informations boursières	stocks: Exemple : stock:yahoo trouvera des informations boursières relatives à la société Yahoo!.
Recherche dans le titre (1)	allintitle: Exemple : allintitle:abondance moteur trouvera les pages qui contiennent les mots abondance ET moteur dans leur titre.
Recherche dans le titre (2)	intitle: Exemple : intitle:abondance moteur trouvera les pages qui contiennent le mot abondance dans leur titre et le mot moteur dans la page.
Recherche dans l'url (1)	allinurl: Exemple : allinurl:abondance moteur trouvera les pages qui contiennent les mots abondance ET moteur dans leur adresse (url).
Recherche dans l'url (2)	inurl: Exemple : inurl:abondance moteur trouvera les pages qui contiennent le mot abondance dans leur url et le mot moteur dans la page.
Recherche dans le texte	intext: Exemple : intext:IBM trouvera les pages qui contiennent le mot IBM dans le texte de la page et pas obligatoirement dans le titre, l'en-tête et le texte des liens.
Recherche avancée	Advanced Search, Recherche avancée Recherche sur le format de fichiers, sur la date de mise à jour, sur la langue, etc.

Référencement

Vérification de la présence de vos pages dans l'index	Saisissez la requête : site:www.monsite.com qui vous indiquera la liste des pages de votre site présentes dans l'index de Google.
Nom et adresse de la fonction de soumission volontaire de pages	Add/Update your URL, Ajouter l'url de votre site Il est recommandé d'inscrire son site sur l' Open Directory et/ou sur Yahoo! avant de le soumettre à Google, cela accélère grandement la prise en compte par le moteur. Plus vous aurez de liens vers votre site et plus l'indexation sur Google sera également rapide.
Délai de prise en compte de la page lors d'une soumission manuelle	2 à 4 semaines (variable en fonction du nombre de liens pointant vers votre site)

Modalités d'exclusion des pages

Fichier robots.txt	Oui : pris en compte
Balise META robots	Oui : prise en compte



Toute l'info et l'actu sur les annuaires et moteurs de recherche : Recherche d'information et référencement

- Vous débutez ?**
- Actualité
- Blog
- L'actu sur votre site
- Dossiers / Articles
- Méthodologies
- Outils de recherche
- Audits
- Forums / Chat
- Lettres d'information
- Études
- FAQ
- Livres
- Emploi
- Tribune
- Humour
- Liens
- Offre commerciale
- Boutique en ligne
- Zone Abonnés**
- Réseau Abondance

[Accueil](#) > [Outils de recherche](#) >

A la une de l'actu

- ▶ Webmasters, **installez les titres de l'actualité** sur votre site !
- ▶ **Toute l'actualité de la semaine**

Exalead



Présentation générale

Nom	Exalead
URL	http://www.exalead.com/
Date de lancement	2002
Propriétaire	Exalead
Taille de l'index	8 milliards de pages
Délai de rafraîchissement de l'index	4 semaines (variable)
Nom du spider	Exalead
Technologie utilisée par les portails :	Exalead, AOL.FR

Syntaxe d'interrogation

Minuscules / majuscules	Idem : Exalead ne tient pas compte de la casse des lettres. Exemple : ibm trouve ibm , IBM ou Ibm . Ibm trouve Ibm , ibm ou IBM . IBM trouve IBM , ibm ou Ibm .
Lettres accentuées	Pas d'influence. Exemple : electricite trouve electricite , électricité ou electricité . électricité trouve électricité , electricite ou electricité .
Ordre des mots	Pas d'influence : paris dakar donne un résultat identique à dakar paris .
OU	OR. L'opérateur OPT également de proposer un mot qui ne sera pas obligatoirement dans les pages de résultats identifiées par le moteur.
ET	Opérateur par défaut Exemple : moteur recherche recherche les pages qui contiennent à la fois moteur ET recherche . Il est également possible d'utiliser l'opérateur AND ou le signe +
SAUF	- Exemple : moteur -automobile recherche les pages qui contiennent moteur mais qui ne contiennent pas automobile . Il est également possible d'utiliser l'opérateur NOT
Expressions	Guillemets. Exemple : "moteur de recherche" recherche cette expression, les trois mots les uns à côté des autres et dans cet ordre.
Troncature	Astérisque (*) L'astérisque (*) permet de rechercher les mots ayant la même racine (lemmatisation).
Recherche dans le titre des pages	title: Exemple : title:abondance moteur trouvera les pages qui contiennent le mot abondance dans leur titre et le mot moteur dans la page.
Recherche dans un nom de domaine donné	site: Exemple : +moteur +site:honda.com recherchera les documents qui contiennent le mot moteur et qui sont disponibles sur le site honda.com .
Recherche sur les langues	language: Exemple : engine language:fr trouvera les pages qui contiennent le mot engine mais qui sont disponibles en français.
Recherche avancée	Advanced Search

Référencement

Vérification de la présence	Saisissez la requête : site:www.monsite.com qui vous indiquera la liste des pages de votre
------------------------------------	---

Chercher sur le Web

avec  mozbot :

Chercher sur :

- Web international
- Web francophone

REFERENCEMENT 2.0

Olivier Andrieu

Le livre sur le référencement de sites web



de vos pages dans l'index	indiquera la liste des pages de votre site présentes dans l'index du moteur.
Nom et adresse de la fonction de soumission volontaire de pages	Submit a site
Délai de prise en compte de la page lors d'une soumission manuelle	Variable

Modalités d'exclusion des pages	
Fichier robots.txt	Oui : pris en compte
Balise META robots	Oui : prise en compte

Un site du **Réseau Abondance** (english version : **Search Engine Tools**) :

Information : Abondance actu référencement - Forums Abondance - Baromètre Référencement - Boutique Abondance - Livre Référencement - Livre Google
Outils : Outiref - Visiref - Keyword Search Engine - Spider Simulator - Soumission-Manuelle - Flash Moteurs - Moklic
Divers / Jeux : Googlefight - LotoFoot Googlefight - Habitants - Googland - Breitling Jet Team - Meeting aérien Lens - Maison en Provence - Grifil.com
Moteurs de recherche : Mozbot.fr - Mozbot.com - Mozbot.co.uk - Goossip - Motref - Rugby engine - Search In The Matrix - Grifil.net - Biomalin - Foot Engine - SearchWings

Recommander ce site à des amis | Bienvenue | Pourquoi Abondance ? | Contact | Crédits | Rechercher sur le site
 Applicatifs réalisés par Brioude Internet et Raynette | Site hébergé par GPL Service | Publicité sur le site | Plan du site



Les 6 étapes (résumé)



Les 6 étapes d'un projet de recherche d'information ©

Pédagogie du projet
1996-2009

[deutsch](#) | [english](#) | [español](#)

Les étapes présentées ci-dessous forment un tout. Dans une situation concrète de résolution de problème d'information, elles peuvent être abordées l'une après l'autre ou plusieurs à la fois, avec des répétitions et des retours en arrière, selon les habitudes de travail et le rythme de chaque apprenant.

<p><u>Étape 1</u> Je cerne le sujet</p>	<ul style="list-style-type: none"> ● Je comprends la nature, les objectifs et l'étendue du travail à exécuter ● Je précise la question de recherche ● Je me lance dans un remue-méninges ● Je dégage les idées importantes et les mots-clés ● Je détermine l'angle sous lequel je vais aborder mon sujet ● Je formule mon idée directrice ● J'esquisse un plan provisoire ● Je réfléchis aux ressources dont j'aurai besoin ● Je planifie mon travail 		
<p><u>Étape 2</u> J'interroge des sources d'information</p>	<ul style="list-style-type: none"> ● Je choisis les ressources dans lesquelles faire ma recherche ● Je construis des équations de recherche à partir de mes mots-clés ● J'amorce la recherche à l'aide de mes mots-clés à la bibliothèque, sur Internet ● Je juge de l'efficacité de mes équations de recherche et je les ajuste au besoin ● Je repère les documents et les ressources qui me semblent pertinents ● J'échange et je partage des idées avec différents interlocuteurs 		
<p><u>Étape 3</u> Je sélectionne des documents</p>	<ul style="list-style-type: none"> ● J'organise les documents que j'ai retenus ● J'évalue la qualité de l'information selon des critères posés ● J'examine différents points de vue ● J'identifie les éléments d'information nécessaires à mon travail ● Je note la référence et les caractéristiques des documents retenus 		
<p><u>Étape 4</u> J'extrais des informations</p>	<ul style="list-style-type: none"> ● Je lis, j'écoute, je visionne attentivement et je prends des notes ● J'indique mes sources d'information et je respecte les règles de la vie privée ● J'organise mes notes de manière cohérente et j'ajuste mon plan provisoire ● Je fais le point sur ma progression et je reprends une étape au besoin 		
<p><u>Étape 5</u> Je traite les informations</p>	<ul style="list-style-type: none"> ● J'analyse les informations prélevées au regard de la question de recherche ● J'envisage différentes solutions permettant de traiter mon sujet ● Je synthétise les idées et les informations provenant de sources variées ● Je confronte des opinions et je construis la mienne selon l'objectif du travail ● Je confirme ou je reformule mon idée directrice ● Je réfléchis à l'organisation de l'ensemble de ma production 		
<p><u>Étape 6</u> Je produis un travail</p>	<ul style="list-style-type: none"> ● Je construis mon travail en fonction de mes objectifs et du public-cible ● J'exprime mon opinion si l'objectif du travail le demande ● Je m'assure que le contenu de ma production est cohérent ● Je révise la qualité de la langue ● Je vérifie si j'ai respecté les consignes ● Je gère le matériel et les appareils dont je peux avoir besoin 		
<p>Retour sur ma démarche</p>	<p>J'utilise les TIC</p>	<p>Nous travaillons en équipe</p>	<p>Guide rapide Sommaire Départ</p>

Cette démarche est utilisée, entre autres, en pédagogie du projet (ou pédagogie par projet ou par projets, pédagogie de projet ou de projets).

Démarche adaptée et mise à jour par [Hélène Guertin](#) à partir de l'ouvrage *La recherche d'information à l'école secondaire : l'enseignant et le bibliothécaire, partenaires de l'élève* (1997) - [Crédits](#)

Note : Le travail d'élaboration de la démarche a bénéficié de l'accès privilégié au document de travail daté de 1996, obtenu avec la permission de Yves Léveillé, dont le titre provisoire était

La recherche d'information à l'école secondaire : un projet de recherche d'information en six étapes.

La présente **version remaniée (2005)** respecte l'esprit du document : [Les compétences transversales](#), dans

Programme de formation de l'école québécoise, enseignement secondaire (2004), chap. 3. Québec : ministère de l'Éducation du Québec. Page consultée le 5 juillet 2005.

[Autres sources consultées](#) pour la mise à jour de 2005.

Avec la collaboration de [Paulette Bernhard](#), professeure honoraire,
École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal, Québec.

[Autres modèles du processus de recherche d'information](#) (site Form@net)

Selon le contexte, "le prof" ou "ton prof" peut être remplacé par "les/tes profs"
et par "le/ton bibliothécaire/enseignant-documentaliste".

L'utilisation du genre masculin a été adoptée afin de faciliter la lecture et n'a aucune intention discriminatoire.



Commentaires, corrections, suggestions

Hélène Guertin : guertinh@videotron.ca



[ACCUEIL](#)



[Les étapes en détail](#)



[HAUT DE PAGE](#)



[GUIDE RAPIDE](#)

Site [Chercher pour trouver : L'espace des élèves](#)

<http://www.ebsi.umontreal.ca/jetrouve/projet/index.htm> - 1996-2009

7. Les références sont citées

Commentaire :

Graphisme | Images | Son | Vidéo qui ajoutent de la valeur au site

1. Le graphisme du site est attrayant
2. Le texte est lisible sans effort
3. Les pages Web sont imprimées correctement
4. Les images sont utiles | Il n'y en a pas
5. Les séquences sonores sont utiles | Il n'y en a pas
6. Les séquences vidéos sont utiles | Il n'y en a pas

Oui	Moyen- nement	Non	Je ne sais pas
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Commentaire :

Utilité du site pour mon projet

1. Ce site est important pour mon projet
2. Ce site couvre l'ensemble de mon sujet
3. **J'accorderais la cote suivante à cette page Web :**

Oui	Moyen- nement	Non	Je ne sais pas
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Excellent Très bon Bon Passable
Sans intérêt pour mon projet

Commentaire :

Principaux documents consultés

1 - **IMAGE** (retouchée) : © ArtToday Inc. « 830676.jpg », sur le site clipart.com
Cette image n'est plus disponible à son adresse d'origine.

 [ACCUEIL](#)  [SECTION INTERNET](#)  [HAUT DE PAGE](#)

Site [Chercher pour trouver : L'espace des élèves](#)

<http://www.ebsi.umontreal.ca/jetrouve/internet/evalsite.htm> - Page mise à jour le 15 février 2009



Section Internet

J'évalue le contenu d'une page Web

Tout internaute peut créer des pages Web et les rendre disponible sur le Net. Comment savoir si le contenu d'une page Web est valable pour ton projet ? Voici quelques éléments d'évaluation.



Voir aussi : Évaluation d'un site Web

Nom de la page :
Type d'information ou de service :
Public-cible : Langue :
Adresse URL :
Page évaluée par : Date :

QUESTIONS

RÉPONSES

Auteur de la page

(les renseignements sur l'auteur sont généralement sur la page d'accueil)

- 1. L'auteur de la page est identifié
2. La fonction et/ou les compétences de l'auteur sont mentionnées
3. L'auteur est affilié à un organisme officiel à une entreprise privée l'auteur est un particulier
4. La date de publication / mise à jour de la page est récente
5. L'adresse de courriel de l'auteur est disponible

Table with 3 columns: Oui, Non, Je ne sais pas. Rows 1-5.

Commentaire :

Contenu de la page

- 1. L'information correspond à mon niveau d'études
2. L'information est présentée de manière méthodique
3. L'information est rédigée ou résumée clairement
4. L'information semble exacte et objective
5. L'information contredit celle que j'ai trouvée ailleurs (ceci n'est pas un point négatif).
6. L'information est rédigée en bon français

Table with 4 columns: Oui, Moyennement, Non, Je ne sais pas. Rows 1-6.

QUESTIONS

RÉPONSES

Contenu de la page (suite)

- 7. Les références sont citées
8. Les images sont utiles | Il n'y en a pas
9. Les séquences sonores et/ou vidéo ajoutent de la valeur au texte Il n'y en a pas
10. La publicité nuit à la compréhension de l'information Il n'y en a pas

Table with 4 columns: Oui, Moyennement, Non, Je ne sais pas. Rows 7-10.

Commentaire :

Importance de cette page Web pour mon projet

1. Cette page est importante pour mon projet

Oui	Moyen- nement	Non	Je ne sais pas
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. **J'accorderais la cote suivante à cette page :**

Excellent Très bon Bon Passable
Sans intérêt pour mon projet

Commentaire :

[Principaux documents consultés](#)

[1](#) - **IMAGE** (retouchée) : © ArtToday Inc. « 2907496.jpg », sur le site clipart.com
Cette image n'est plus disponible à son adresse d'origine

[🏠 ACCUEIL](#) [◀ SECTION INTERNET](#) [▶ HAUT DE PAGE](#)

Site [Chercher pour trouver : L'espace des élèves](#)

<http://www.ebsi.umontreal.ca/jetrouve/internet/evalpage.htm> - Page mise à jour le 15 février 2009



REPERE

Ressources Electroniques
Pour les Etudiants,
la Recherche et l'Enseignement

<http://reperre.enssib.fr>

2008

Techniques d'accès aux ressources électroniques et à Internet

Auteurs :

Marie-Françoise DEFOSSE
Professeur associée - ens sib

Elisabeth NOËL
Conservateur des bibliothèques - ens sib



La connaissance des techniques de l'Internet est utile pour mieux comprendre les résultats des recherches d'informations, pour l'évaluation des sources d'informations et pour leur utilisation.

Dispositifs techniques d'accès à l'information

Internet, qu'est-ce-que c'est ?

INTERNET (INTERconnected NETWORK) est né en 1969 du besoin conjoint des chercheurs américains et de l'armée de communiquer via des ordinateurs. Au début des années 90, l'arrivée des ordinateurs individuels a démocratisé l'accès au réseau sur le lieu de travail, la maison et les lieux publics.

Internet est devenu un instrument de communication ouvert à tous pour communiquer, échanger, consulter des documents et aussi en publier. Tout le monde peut être à la fois lecteur et fournisseur d'informations, d'autant que le web 2.0 facilite la production. Mais rien n'authentifie les informations, qui peuvent être falsifiées (problèmes de date, de mise à jour...) ou erronées (**hoax** : fausses informations, canulars), d'où l'importance de l'évaluation des sites Internet.

Personne ne connaît exactement l'étendue d'Internet, son contenu demeure éphémère et volatile. L'archivage du web

n'est pas réellement assuré, malgré des initiatives intéressantes (Internet Archive, <http://www.archive.org/> ou le dépôt légal Internet, à la Bibliothèque Nationale de France).

Aussi, il est toujours possible d'intercepter les paquets d'information qui circulent dans le réseau. D'où aussi des problèmes de sécurité des transactions commerciales ou de confidentialité sur les sites non sécurisés. Un site sécurisé se reconnaît par la présence d'un cadenas fermé ou d'une clé intacte (non brisée) dans le coin inférieur droit ou gauche de l'écran, selon le navigateur utilisé. De plus, l'adresse URL d'un site sécurisé commence habituellement par « **https** » ou « **shttp** » au lieu de « **http** ».

Enfin, le téléchargement de logiciels entraîne toujours le risque de rapatrier des virus ! Il est donc conseillé de vérifier chaque logiciel ou document récupéré par un anti-virus à jour, les meilleurs étant gratuits. Il est aussi utile de se prémunir contre les *spywares*, logiciels espions qui

- LEVINE John R. *Internet pour les nuls*. Paris : Éditions Générales First , 2008.
- VIRIEUX Françoise. *Comment marche Internet ?* Paris : le pommier, 2004 (Les petites pommes du savoir).

fournissent aux sociétés qui les diffusent des informations sur les habitudes des internautes, sans que ceux-ci en soient informés.

Connexion à un réseau

Le protocole de base : TCP-IP

Du point de vue technique, Internet est un terme générique qui désigne la multitude des réseaux d'ordinateurs qui communiquent entre eux, au niveau mondial, grâce à un protocole d'échange de données standard : TCP/IP (Transmission Control Protocol/Internet Protocol – se prononce T-C-P-I-P), indépendamment des types d'ordinateurs utilisés (Mac, PC).

TCP/IP est basé sur le repérage de chaque ordinateur par une adresse appelée **adresse IP**, qui permet d'acheminer les données à la bonne adresse. Par exemple, 194.153.205.26 est une adresse TCP/IP donnée sous sa forme technique. Ce sont ces adresses que connaissent les ordinateurs pour communiquer entre eux.

Les noms de domaines

Les utilisateurs ne veulent pas travailler avec des adresses numériques mais avec des noms plus explicites. Le protocole TCP/IP a donc permis d'associer des noms en langage courant aux adresses numériques, grâce à un système appelé DNS (*Domain Name System*). Le **nom de domaine** est le nom à deux composantes, dont la première partie correspond au nom de l'organisation ou de l'entreprise, la deuxième à la classification de domaine (.fr, .com, ...) – c'est l'extension.

Exemple : dans l'adresse <http://www.enssib.fr>, « enssib.fr » est le nom de domaine : enssib est le nom de l'organisation, .fr correspond à la classification. Les « www » ne correspondent à rien de significatif, ils peuvent être remplacés par d'autres lettres, si elles sont déclarées. Ainsi, « formist.enssib.fr » correspond à un site particulier dépendant du domaine enssib.fr.

Chaque machine d'un même domaine est appelée **hôte**. Le nom d'hôte qui lui est attribué doit être unique dans le domaine considéré.

Syntaxe d'une URL

Une URL (*Uniform Resource Locator*) est un format de nommage universel pour désigner une ressource sur Internet. Il s'agit d'une chaîne de caractères qui se compose de quatre parties :

- 1- Le nom du **protocole** : le langage utilisé pour communiquer sur le réseau. Ce n'est pas toujours « http ».
- 2- L'adresse du serveur : le **nom de domaine** de l'ordinateur hébergeant la ressource demandée. Il est possible d'utiliser l'adresse IP du serveur, ce qui est moins pratique.
- 3- Le numéro de port : facultatif, ce numéro associé à un service permet au serveur de savoir quel type de ressource est demandé.
- 4- Le **chemin d'accès** à la ressource : cette dernière partie indique au serveur l'emplacement où la ressource est située, c'est-à-dire le chemin (répertoire et sous-répertoires) et le nom du fichier demandé.

Une URL a donc la structure suivante :

Protocole	Adresse du serveur	Port (facultatif)	Chemin /fichier
http://	www.enssib.fr		/glossaire/glossaire.htm

Exemples : http://www.enssib.fr/bibliotheque/cadre_catalogues.html
<news://fr.doc.biblio>

Les réseaux de télécommunications

En France, le réseau **Renater** (REseau NAional de télécommunication pour la Technologie, l'Enseignement et la Recherche), géré par un groupement d'intérêt public, interconnecte les différentes universités et organismes de recherche français au niveau national, avec les autres réseaux de la recherche au niveau européen et avec l'Internet, en général.

D'autres réseaux de communication permettent également une connexion à Internet : en France, il s'agit de Transpac et Numéris. Les particuliers avec un ordinateur équipé d'un modem peuvent ainsi être connectés à Internet via le réseau téléphonique commuté (ou Numéris), par l'intermédiaire d'un prestataire qui leur fournit le point d'accès. Le **Fournisseur d'accès à Internet** (FAI, "provider" en anglais) permet de se connecter à Internet via un modem, par une ligne téléphonique, par l'ADSL (*Asymmetric Digital Subscriber Line*), par le câble ou par le Wi-Fi (*Wireless Fidelity* - réseau local sans fil). Il offre différents services : adresses de messagerie, hébergement de pages personnelles, service technique (hot-line)... Son choix doit prendre en compte les besoins, les services offerts, leur qualité et leur coût.

Protocoles de connexion

Le réseau offre différents types de services :

SMTP - Le courrier électronique (courriel, mél ou e-mail)

Le protocole **SMTP** (*Simple Mail Transfer Protocol*) est le protocole standard du courrier électronique, de la messagerie individuelle, des listes de discussion, qui permet de transférer le courrier d'un serveur à un autre.

Des logiciels installés sur l'ordinateur (comme Thunderbird, Eudora ou Outlook)

aident à gérer facilement son courrier grâce à des interfaces conviviales. Les messages reçus restent stockés sur le serveur jusqu'à leur transfert sur l'ordinateur client pour consultation. La messagerie est aussi consultable directement en ligne, sur le site du fournisseur d'accès, à travers des services webmail qui permettent de gérer tous les messages directement sur le serveur. Si la consultation nécessite d'être connecté, elle permet cependant de lire et d'envoyer des mails depuis n'importe où, avec un même compte mail.

Pour correspondre avec une personne où qu'elle se trouve dans le monde, il suffit qu'expéditeur et destinataire possèdent une adresse électronique qui se présente sous la forme :

nom_usager@adresse_organisation.extension

Il n'existe pas d'annuaire mondial des adresses Internet ; même si plusieurs services permettent des recherches, le résultat est rarement performant.

FTP - Le transfert de fichier

Le protocole **FTP** (*File Transfer Protocol*) est un protocole de transfert de fichier, qui définit la façon selon laquelle des données doivent être transférées sur un réseau TCP/IP. Il permet :

- un partage de fichiers entre machine distante,
- une indépendance aux systèmes de fichiers des machines clientes et serveur,
- le transfert efficace des données.

Ce protocole permet d'accéder à un serveur distant et de transférer des fichiers contenant des textes, des images, du son ou des logiciels. Il nécessite de disposer d'un programme de décompression (winzip). Il peut être utilisé via un navigateur ou par un logiciel spécifique.

La syntaxe est la suivante :

ftp://adresse_serveur/répertoire/

sous-répertoire/fichier

Ex : ftp://ftp.inria.fr/

NNTP - Forums de discussion

Le protocole des **forums** est le *Network News Transfer Protocol*. Les **news** ou **newsgroups** sont consultables sur les serveurs Usenet (news : nom_du_groupe), à travers un logiciel spécifique ou intégré, et divisés en grandes catégories selon des thèmes définis comme :

Préfixe	Contenu
Alt	discussions "alternatives" abordant les sujets les plus variés
Bionet	recherche en sciences de la vie, biologie
Comp	tous les domaines de l'informatique
Fr	groupes francophones, messages en français
Rec	loisirs, groupes des passionnés

Sci	recherche en sciences exactes
Soc	recherche en sciences humaines, sociales

Les adresses sont structurées dans le sens inverse des URL.

Exemple : fr.bio.medecine correspond au domaine France, au thème biologie, au groupe Médecine.

HTTP - L'accès aux serveurs d'informations WWW

Le protocole HTTP (*Hypertext Transfer Protocol*) est le protocole de communication entre un serveur WWW, offrant des contenus multimédias reliés par des liens hypertexte, et un logiciel client, le navigateur (Internet Explorer ou Firefox par exemple). Ce service Internet, qui permet de naviguer à travers des pages web, est généralement appelé "web".

Nature du web

Le **world wide web** est un service récent d'Internet. C'est en 1990 que des informaticiens du CERN (Centre Européen de Recherche Nucléaire) proposent cette application destinée à la publication et la diffusion de documents multimédia sur le réseau, au sein de la communauté internationale des physiciens.

Le web s'est imposé grâce à ses qualités :

- le recours au multimédia, qui mixe le texte, les images fixes et animées, le son ;
- l'hypertexte, qui permet une navigation intra et inter-documents et a permis de tisser une « toile » mondiale.

Modalités de connexion au web

Internet privilégie l'environnement **client-serveur** : l'utilisateur travaille donc depuis son propre ordinateur, sur lequel ont été transférés les fichiers demandés au site consulté. Quand nécessaire, des échanges

se font entre l'ordinateur individuel (le client) et le serveur. L'affichage se fait grâce au langage HTML (*HyperText Markup Language*) qui permet d'indiquer au client (ici, le navigateur) comment interpréter les données. L'option Affichage/Source dans le menu du navigateur permet de voir une page telle qu'elle est codée en HTML.

Certains documents nécessitent l'activation de **plugins**, logiciels qui s'exécutent à l'intérieur du navigateur : Realplayer, Shockwave..., pour le son et les animations. Ces logiciels sont généralement gratuits.

Le web en 2008

La facilité de production de pages HTML et la popularité croissante des blogs, l'existence de nombreux sites d'hébergement gratuits, les faibles coûts des serveurs ont conduit à un développement quantitatif

énorme du nombre de pages HTML depuis 1995.

Le web est souvent divisé en trois parties : le web public, libre d'accès, le web privé, sur mot de passe et le web provisoire, en construction. Les moteurs de recherche indexent en grande partie le web public, mais ignorent le reste, qui demeure invisible à leurs yeux.

Web statique, web dynamique

Le standard établi pour la diffusion de documents sur le web est le langage HTML, qui permet de définir la présentation du document ainsi que les liens hypertextes vers d'autres documents à l'aide de balises de formatage. La description du contenu a été prévue dans les balises de métadonnées (metadata : données sur les données), étendues avec le Dublin Core (*Dublin Core Metadata Initiative*, <http://dublincore.org/>). Cependant pour diverses raisons, depuis 2001 les moteurs de recherche prennent peu en compte ces éléments, qui restent utiles dans le cadre d'un travail en réseau documentaire ou en intranet avec utilisation d'un moteur d'indexation.

Les documents fixes produits à l'aide d'un éditeur de texte ou d'un éditeur HTML sont stockés en tant que fichiers texte (extension HTM, HTML, SHTML...) sur le disque dur du serveur web qui les diffuse. Ils constituent des pages HTML statiques.

Cependant, il est « plus facile » de gérer une base d'informations que de modifier des centaines de textes HTML. D'où le développement de bases de données, qui diffusent l'information à la demande, suite à l'interrogation de la base de données par l'intermédiaire d'un formulaire : il s'agit alors de **pages HTML dynamiques**.

Le web invisible

Le web invisible est défini comme l'ensemble des documents diffusés sur le web, sans être indexés par les moteurs de recherche traditionnels. Plusieurs raisons dues aux méthodes de référencement dans les outils de recherche l'expliquent. Les moteurs de recherche envoient des robots balayer le Web pour alimenter en données leur index, mais certaines pages web ne peuvent pas être indexées : mot de passe demandé (sites commerciaux permettant le paiement en ligne par carte bancaire, Intranet), niveau de profondeur d'indexation ou de profondeur de contenu indexé, pages vers lesquelles ne pointe aucun lien, formats de fichiers pas reconnus, indexation difficile des pages web dynamiques, mauvais référencement ou volonté des auteurs et responsables de sites, etc.

Le web invisible regroupe des informations en général intéressantes, spécialisées, appartenant à la littérature grise, aux archives ouvertes et aux bases de données.

- **AFNIC. Chartes de nommage [en ligne].**

<http://www.afnic.fr/obtenir/chartes>

- **Comment ça marche. Les adresses IP [en ligne].**

<http://www.commentcamarche.net/internet/ip.php3>

- **Internet Archive [en ligne].**

<http://www.archive.org/>

- **USENET fr. Liste des groupes FR [en ligne].**

<http://www.usenet-fr.net/liste-groupes.html>



Gestion des signets, favoris ou bookmarks

Auteurs :

Gabriel GALLEZOT

Maitre de conférence – URFIST de Nice

Michel ROLAND

Conservateur des bibliothèques – URFIST de Nice

- 86** Utilisation d'outil intégré au navigateur
- 86** Les gestionnaires de favoris en ligne
- 88** La gestion des références en ligne



Les gestionnaires de favoris permettent à l'utilisateur du web de mémoriser ses "trouvailles", soit dans son navigateur, soit en ligne, pour mieux y retourner par la suite.

Utilisation d'outil intégré au navigateur

L'URL de la page web peut être simplement mémorisée dans le navigateur (fonction : ajouter un signet, favori, bookmark, marque-page), éventuellement dans des dossiers personnalisables, et rendue accessible *via* le menu déroulant. Cependant, il devient vite difficile

de retrouver facilement des adresses, en raison de leur grand nombre et du manque de possibilité de description fine. Cette fonction reste utile pour garder « sous la main », dans la barre personnelle, les signets les plus fréquemment utilisés, les outils en particulier.

Les gestionnaires de favoris en ligne

Les outils sociaux du web 2.0 permettent d'accéder de partout (nomadisme) aux liens enregistrés, d'étiqueter (*tagging*) les signets dans des dossiers et de les partager, en profitant du travail de sélection fait par d'autres utilisateurs de l'outil partageant des intérêts semblables.

Del.icio.us

Del.icio.us est le plus populaire des outils de gestion de signets en ligne, mais pas le seul. Il nécessite de se créer d'abord un compte, pour accéder par la suite au service après identification. *Del.icio.us* peut ensuite être intégré dans le navigateur, soit par extension ou module complémentaire, soit par l'ajout d'un *book-*

marklet, qui lance une mini-application pour effectuer certaines opérations. Deux boutons apparaissent alors dans la barre de navigation, l'un pour mémoriser l'URL de la page web, l'autre pour ouvrir la page web du compte *Del.icio.us* de l'utilisateur. Enfin, il reste toujours possible de mémoriser l'URL d'une page directement sur *Del.icio.us*.

L'enregistrement d'une URL permet d'ajouter des éléments d'information, zone de note pour décrire le contenu de la page, et **étiquette (ou tag)**, qui ressemble à un mot-clé, mais donne en priorité un sens parlant à l'utilisateur avant de décrire objectivement le contenu.

Save a Bookmark

delicious Save a Bookmark Signed in as urfirst_pac

URL do not share

TITLE

NOTES 255 chars

TAGS ?

Separate Tags with spaces, E.g., hotels bargains newyork (not new york)

Recommended Tags: click to add from your existing tags
art videos

Network Tags: click to share this link with your network
for:bibliothecaire for:cercamon for:chimist for:comedd for:gaga68 for:n_anna for:wiki_ao

Popular Tags: click to add from popular tags on
art graffiti culture design streetart video graphics green

Save Cancel

Formulaire de saisie des informations sur Del.icio.us

La liste d'étiquettes déjà attribuées par d'autres à la page web est proposée, ainsi que la possibilité de rendre le signet public ou privé (*do not share*). Cette possibilité doit cependant être activée. La page d'accueil d'un compte liste les signets mémorisés en ordre chronolo-

gique inverse (les derniers arrivent en haut de page), ainsi que les étiquettes associées (en liste ou « nuage de tags »), le nombre de personnes qui ont sauvegardé ce signet sur del.icio.us (indice de popularité) et la date de sauvegarde.

del.icio.us / bibliothecaire /

your bookmarks | your network | subscriptions | links for you | post

logged in as bibliothecaire | settings | logout | help

All your items (2396) search

« earlier | later » page 2 of 24

ATHINAIS HOTEL edit | share / delete
to hater hotels for:cercamon ... saved by 1 other person ... not shared ... on April 26

title **affordance.info: Faut-il stabiliser Wikipédia ?**

url http://affordance.typepad.com/mon_weblog/2008/03/faut-il-stabili.html

notes

tags **wikipedia**

save cancel do not share full-screen edit

affordance.info: Mon historique de navigation n'appartient (et le reste aussi) edit | delete
to affordance.info ... saved by 1 other person ... on April 26

affordance.info: Le plug-in de la redocumentarisation, edit | delete
to wordpress blog for:compartiment bright ... saved by 1 other person ... on April 26

Word macros - Meta edit | delete
to wiki markup ... on April 25

Wiki Markup :: start edit | delete
to wiki markup ... on April 25

MDC-Wiki Markup Reference - MDC edit | delete
to wiki markup ... saved by 13 other people ... on April 25

Column Two: Wiki markup has no future edit | delete
to wiki markup ... saved by 7 other people ... on April 25

- collecting **wikigraphy** bookmarking access.us feeds reference_manager referencemanager
- referencing networks social news
- creating_content biblioblog bibliowiki sau blog blogging blogs desktop CMS content_management ESP engine libraries microblogging platform wordpress wiki wikis wikimedia wikis wiki_engine wordpress
- open_access archive archiving CHRIS communication browser CPU creative commons digital libraries digital_newsletters e_journal e-journals e-print Elsevier EPST EurOpenHolar evaluation institutional_repositories interoperability JISC JISC-UK-CHN OAI-PMH OAI open_access open_access open_sciences open_sciences production_scientific publication_scientific res ranking taxonomy research scholars_scholarship_sciences_2.0 news Strength
- people Barack-Obama large Olivier-Jean-Louis-Jay-Denis-Lafont-Dorval-Felakkael-Henric-huguelle_rigot Huang lich jerry_wells James-Beckwith-Lorenson-Loring-Markus-Andreas-Oppe-Jaime-Skvarna-Olivier-LaDuff-Steve-John-Dawson-Coburn
- reading arxivmeds_personal books e-book e-books e-paper e-readers lich library reading
- signetsBPFirefox bkmk CIL Cerco vertebrales chercheurs DBP de communication thesaurus thesauri thesauribookmarking thesauri

Page d'accueil d'un compte Del.icio.us

Le fait de voir qui a enregistré le même signet, puis les signets de cet internaute, de trouver de nouveaux signets associés à un tag utilisé correspond à du *social bookmarking*. D'autres fonctions, de recherche dans *Del.icio.us*, d'abonnements aux signets d'un autre internaute ou relatifs à un tag défini, de création de fils de syndication (RSS), de partage des signets recueillis enrichissent sans cesse l'utilisation de *Del.icio.us*.

Diigo

Diigo est un autre outil de gestion de signets en ligne, plus complexe et plus puissant, qui offre des fonctions avan-

cées permettant d'interagir avec d'autres applications web. Il permet, outre la mémorisation de base de type *Del.icio.us*, de garder des extraits et de rajouter des annotations, comme lors d'une lecture « studieuse », crayon à la main. Ces annotations peuvent aussi être vues par d'autres, si elles sont rendues publiques.

Les outils de gestion de signets, en ligne ou non, possèdent généralement des fonctions d'import et d'export permettant des échanges ou des transferts faciles.

The screenshot shows the Diigo web interface. At the top, there is a navigation bar with links like Home, Message (5), Friends, Meet People, Communities, Discover, and a search bar. Below this is a menu with options like My Dashboard, My Bookmarks, My Lists, My Tags, My Sites, and My Groups. The main content area is titled 'My Bookmarks' and includes a filter box and a list of bookmark entries. One entry is expanded, showing a bookmark titled 'Aide de Windows Vista : Que sont les favoris Internet Explorer ?' with tags like 'IE', 'bookmarking', and 'about 22 hours ago'. Another entry is titled 'OPACs à la loupe : Marlene's corner' with tags 'opac', 'library_2.0', and 'web'. A third entry is 'The Library in the New Age - The New York Review of Books' with tags 'book', 'readings', 'library', and 'starbon'. To the right of the bookmark list is a 'My Recent Tags (50)' cloud containing various tags such as 'architecture', 'attention', 'bibliography', 'blogging', 'book', 'bookmarking', 'copyright', 'decline', 'desktop_tools', 'digital_natives', 'droit_d_auteur', 'droits', 'e_readers', 'epst', 'facebook', 'flickr', 'flickr_export', 'folksonomy', 'fonctionnariat', 'france', 'identity', 'intellectual_property', 'libraries', 'lifestreaming', 'literacy', 'metadata', 'networks', 'non-modern', 'opensource', 'reading', 'reading_2.0', and 'reference'.

Signets sur *Diigo*

La gestion des références en ligne

La popularité et l'efficacité d'un outil comme *Del.icio.us* a inspiré le développement d'outils de même type dans les outils de gestion des références bibliographiques. En effet, l'importance acquise

par Internet pour le monde scientifique a fait de l'accès en ligne une qualité déterminante pour l'utilisateur. Cependant la gestion des références scientifiques suppose l'importation dans l'outil de méta-

données plus riches ; la capacité de l'outil à les récolter automatiquement est un critère de choix essentiel. *CiteULike* est un bon exemple

Ces deux types d'outils croisent certains critères : la nature de l'objet mémorisé

(page/signets, ou référence bibliographique), la localisation de l'outil (local en mode client, hébergé en ligne par un fournisseur ou encore installé sur un serveur local), enfin le coût, les logiciels bibliographiques pouvant être payants.

objets mémorisés ↘ accès à l'outil ↙		SIGNETS (Pages web)	REFERENCES (Documents)	
		Gratuit		Onéreux
LOCAL		Favoris IE Marques-pages Firefox	Bibus Zotero BibDesk	EndNote
EN LIGNE	hébergé	Del.icio.us Ma.gnolia ... Google bookmarks Diigo	CiteULike Zotero 2.0	Refworks EndNote Web
		Bibsonomy Connotea 2collab		
	serveur local	Scuttle	Wikindx (Bibus)	
		Connotea		

Type d'outils de gestion de signets/ références

Ces outils ne sont pas équivalents mais procèdent de la même logique : mémoriser en ligne des références (bibliographiques ou pages web). Certains sont clairement orientés vers les références bibliographiques (*CiteULike*), d'autres "signets" mais permettent, au détour d'un menu dédié, d'intégrer des références bibliographiques

(*Connotea*), d'autres encore affichent simplement leur dualité (Bibsonomy).

CiteULike

Le fonctionnement de *CiteULike* propose trois grands principes d'intégration des références bibliographiques : depuis un Bookmarklet ajouté au navi-

gateur, depuis une liste de sommaires présentée au sein de *CiteULike*, ou encore une intégration manuelle. Des tags peuvent être associés à la des-

cription du document, ainsi que le document source en PDF et la possibilité de partager la référence avec d'autres utilisateurs.



Exemple d'interface de partage de références d'articles (CiteULike)



- URFIST de Nice. *Mémoriser : signets, favoris et références* [en ligne].
http://wiki-urfist.unice.fr/wiki_urfist/index.php/Mémoriser:_signets,_favoris_et_références
- Bibsonomy [en ligne].
<http://www.bibsonomy.org/>
- CiteULike [en ligne].
<http://www.citeulike.org/>
- Connotea [en ligne].
<http://www.connotea.org/>
- Del.icio.us [en ligne].
<http://del.icio.us/>
- Diigo [en ligne].
<http://www.diigo.com/>
- Google Bookmarks [en ligne].
<http://www.google.com/bookmarks/>
- Iterasi [en ligne].
<http://iterasi.com/>
- Ma.gnolia [en ligne].
<http://ma.gnolia.com/>
- Zotero [en ligne].
<http://www.zotero.org>

Les outils de recherche d'information sur le Web

Auteurs :

Marie-Françoise DEFOSSE
Professeur associée – ensib

Élisabeth NOËL
Conservateur des bibliothèques – ensib

- 96** Outils de recherche
- 97** Quel outil choisir ?
- 97** Méthode de tri des résultats
- 98** Un peu de syntaxe



Une multitude d'outils différents composent le paysage de la recherche d'informations sur le web, paysage mouvant (rachats, disparitions temporaires ou définitives, créations incessantes de nouveaux outils) dominé par Google. Seule une bonne connaissance de la nature et du fonctionnement de ces outils, de leur particularité et de la syntaxe des équations de recherche permet d'arriver à des résultats satisfaisants sans y consacrer trop de temps.

Outils de recherche

Les annuaires de recherche

Il s'agit de répertoires (catalogues, guides ou *directories*) de sites classés manuellement en grandes catégories organisées hiérarchiquement. Construits « à la main », après sélection plus ou moins sommaire, ils sont constitués de rubriques (ou catégories), elles-mêmes subdivisées en sous-rubriques. Un annuaire s'utilise par navigation dans les catégories ou par une recherche mots-clés dans les descriptifs des sites. Actuellement, ce modèle a tendance à disparaître, sauf pour les annuaires spécialisés.

Annuaire spécialisés

- CISMef [en ligne].
<<http://www.chu-rouen.fr/cismef/>>
- Intute [en ligne].
<www.intute.ac.uk/>
- Signets de la Bibliothèque Nationale de France [en ligne].
<<http://signets.bnf.fr/>>
- The WWW Virtual Library [en ligne].
<<http://wlib.org/>>

Les moteurs de recherche

Ce sont les outils d'accès au web les plus utilisés. Des robots (appelés crawlers ou spiders) parcourent le Web de lien en lien et indexent automatiquement les documents trouvés. Les moteurs proposent des formulaires de recherche simple et avancée, par mots-clés.

Moteurs de recherche généralistes

- Exalead [en ligne].
<<http://www.exalead.fr/>>
- Google [en ligne].
<<http://www.google.fr/>>
- Yahoo ! [en ligne].
<<http://fr.yahoo.com/>>

Les métamoteurs

Les métamoteurs interrogent plusieurs outils de recherche simultanément, en dédoublonnant les réponses. Sous forme logicielle, installés sur le poste, ils facilitent la veille. Plus exhaustive que sur un

- MANIEZ Dominique. *Les 10 plaies d'Internet*. Paris : Dunod, 2008.
- MESGUICH Véronique, THOMAS Armelle. *Net recherche : Le guide pratique pour mieux trouver l'information utile*. Paris : ADBS, 2007 (Sciences et techniques de l'information).

moteur, la recherche sera néanmoins moins pointue. Ils proposent des spécificités comme la cartographie des résultats (avec Kartoo, <http://www.kartoo.com/>) ou la catégorisation automatique des résultats, c'est-à-dire la création à la volée de dossiers thématiques contenant les résultats trouvés (avec Clusty, <http://clusty.com>).

Métamoteurs

- Copernic (logiciel)
<<http://www.copernic.com/index.html>>

- Ixquick [en ligne].
<<http://eu.ixquick.com/fra/>>
- Metacrawler [en ligne].
<<http://www.metacrawler.com/>>

A noter : les méta-moteurs, qui sont souvent des produits d'appels proposés gratuitement par des éditeurs de logiciels de veille, permettent ainsi de mener sa veille.

Quel outil choisir ?

Le choix de l'outil dépend étroitement du type de recherche à effectuer.

Un annuaire permet de retrouver un ensemble de sites (et non de pages web) centrés sur une thématique spécifique. Ils permettent de débiter une recherche, d'identifier rapidement et précisément des sites de référence et de repérer des ressources du web invisible.

Un moteur de recherche, fonctionnant automatiquement, proposera des résultats plus nombreux mais avec un bruit plus prononcé. La recherche porte sur les pages elles-mêmes : le moteur est à uti-

liser pour la taille de son index, la rapidité de mise à jour de l'index. Enfin, la recherche avancée propose des paramètres très intéressants, permettant de mener des recherches pointues sur un sujet.

Le métamoteur permet d'obtenir un premier aperçu des informations dites « de référence ». Ses résultats dépendent des outils de recherche qu'il utilise.

Les outils de recherche spécialisés s'avèrent souvent être plus efficaces que les outils généralistes.

Méthode de tri des résultats

Les moteurs de recherche complètent la recherche par des techniques de tri des résultats. Plusieurs solutions ont été développées, malheureusement souvent détournées.

Tri par pertinence

Les résultats d'une requête sont affichés selon un ordre déterminé par le calcul d'un score pour chaque réponse. La pertinence est basée sur les 5 facteurs suivants, appliqués aux termes de la question :

- Le poids d'un mot dans un document est

déterminé par sa place dans le document : il est maximum pour le titre et le début du texte ; à l'intérieur du texte, il est plus important si le mot est en majuscule.

- La densité est basée sur la fréquence d'occurrence du terme dans un document, par rapport à la taille du document. Si deux documents contiennent le même nombre d'occurrences, le document le plus petit sera donc favorisé.

- Le poids d'un mot dans la base est basé sur la fréquence d'occurrence du terme dans toute la base de données. Les mots peu fréquents dans le corpus sont favori-

sés. Les mots vides sont soit éliminés, soit sous-évalués.

- La correspondance d'expression est basée sur la similarité entre l'expression de la question et l'expression correspondante dans un document. Un document contenant une expression identique à celle de la question reçoit le poids le plus élevé.
- La relation de proximité est basée sur la proximité des termes de la question entre eux dans le document. Les termes proches sont favorisés.

Tri par popularité

Google a lancé cette méthode, basée sur le nombre de liens pointant sur les pages

(PageRank). La popularité est calculée selon de nombreux critères comme le nombre de fois où le document a été demandé, le nombre de liens hypertextes qui pointent sur le document demandé... Plus un document est demandé, plus il est connu, plus il est populaire et mieux il est placé dans la liste de résultats !

Tri par calcul dynamique de catégories

Cette méthode effectue un classement des documents trouvés dans des dossiers (*clustering*) constitués automatiquement en fonction des réponses. Un dossier peut lui-même être constitué de sous-dossiers.

Un peu de syntaxe

La plupart des outils de recherche sur Internet utilisent la syntaxe suivante, qui pourra être employée sans risque grave. D'autres possibilités de recherche existent, mais varient d'un outil à l'autre.

La recherche avancée permet souvent de poser la question de manière guidée. Le site Abondance propose des fiches pour chaque outil de recherche, présentant les spécificités de syntaxe de chacun.

Opération	Opérateurs	Exemple
Rechercher une expression	" "	"vache folle"
Imposer un terme	+	+vache +folle
Exclure un terme	-	+bretagne -grande

Attention, les signes + et - doivent être collés à gauche du terme concerné. Cette syntaxe correspond en général aux formulaires de recherche simple.

- ANDRIEU Olivier. *Abondance – outils de recherche [en ligne]*.

<http://outils.abondance.com/>

- Bibliothèque de l'UQAM. *Internet [en ligne]*.

<http://www.bibliotheques.uqam.ca/ressources/internet/index.html#mot>

- LARDY Jean-Pierre. *DADI : répertoire des bases de Données gratuites Disponibles sur Internet [en ligne]*.

<http://dadi.enssib.fr/>

