

LE WEB INVISIBLE, cet inconnu

LE WEB INVISIBLE, cet inconnu

DOSSIER DOCUMENTAIRE



Exercice



INVISIBLE WEB

I/ Etre capable de le définir	2
II/ Savoir l'utiliser	3
Annuaire.....	3
Catalogues de bibliothèques	3
Bases de données	4
Portails	4
Moteurs.....	5
III/ Etre incollable	6
Bibliographie.....	6
Sites d'experts.....	6
Newsletters	6

Ce dossier a pour objet de donner des clés pour comprendre ce qu'est le web invisible. Il s'adresse plus particulièrement à des étudiants en documentation, maîtrisant plus ou moins la recherche d'informations sur Internet. Maîtriser les outils du web invisible constitue une valeur ajoutée pour les documentalistes qui se démarquent ainsi de l'internaute lambda.

Définir et expliquer le web invisible :

- ✓ Donner des clés aux étudiants en documentation (type de l'EBD) en leur conseillant des outils spécifiques à la recherche sur le web invisible pour compléter et mettre à jour leurs connaissances générales ainsi que dans leurs secteurs d'activité.

Il ne s'agit pas d'être exhaustif sur le sujet, c'est pourquoi nous avons sélectionné les sites de la manière suivante :

- ✓ Dans la seconde partie, chaque catégorie d'outils contient deux sites complémentaires qui constituent un bon point de départ pour découvrir le web invisible.
- ✓ Une grille d'évaluation présente sous forme de tableaux les critères de sélection les plus importants : auteurs reconnus dans le domaine, contenu riche, classification des résultats, mise à jour, navigation, ergonomie, sites gratuits privilégiés.

- ✓ Le système d'évaluation :   Site à consulter

   Site incontournable

I/ ETRE CAPABLE DE LE DÉFINIR

Le web invisible désigne la partie du web non accessible aux moteurs de recherche classiques. En effet, le problème n'est pas tant la visibilité que l'**accessibilité** par les moteurs.

Le grand dictionnaire terminologique publié par l'Office québécois de la langue française en donne une définition précise : "Les ressources du Web invisible comprennent, entre autres, les sites Web construits autour d'une base de données (interrogeables 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, les intranets et les extranets."

Le web invisible comprend des bases, banques de données et bibliothèques en ligne gratuites ou payantes. Il est aussi appelé web profond (deep web) en opposition au web de surface (surface web).

D'une manière générale, la taille des documents présents sur le web invisible est moins importante que sur le web visible, mais le contenu est plus riche.

- ⇒ Voir le site de [Bright Planet](#), société spécialisée dans la recherche sur le web invisible, qui fournit des données chiffrées (tableaux, graphiques...).

Selon [Chris Sherman et Gary Price](#), deux experts américains des moteurs de recherches, spécialistes du web invisible, il existe quatre catégories de deep web :

- **The Opaque Web** : les pages qui pourraient être indexées par les moteurs mais qui ne le sont pas
- **The Private Web** : les pages web disponibles mais volontairement exclues par les webmasters protégées par des mots de passe
- **The Proprietary web** : pages seulement accessibles pour les personnes qui s'identifient
- **The Truly Invisible Web** : contenu qui ne peut être indexé pour des raisons techniques. Ex : format inconnu par le moteur, pages générées dynamiquement.

- ⇒ Voir le site de [Christophe Asselin](#), expert en intelligence économique et veille stratégique, qui recense dans un très bon dossier sur le sujet un grand nombre d'outils répertoriés en catégories.

II/ SAVOIR L'UTILISER

ANNUAIRES

Ils sont conseillés pour commencer une recherche sur le web invisible. Ce sont des listes de sites web classés et **hiérarchisés** dans des catégories thématiques. Ce travail de référencement est effectué par des **experts** sélectionnés pour leur compétence dans le secteur ou professionnels de l'information-documentation qui jugent et classent ces sites périodiquement.

	<u>DMOZ</u> <i>Anglophone</i> 	<u>Librarian's index to the Internet</u> <i>Anglophone</i> 
Auteurs, éditeurs	Open Directory Project. Experts bénévoles par domaine.	Bibliothécaires universitaires américains.
Contenu	Annuaire collaboratif le plus riche du web, plus de 3 millions de sites répertoriés. Possibilité d'avoir des sélections de réponses en français. Base de répertoire de Google.	Plus de 11000 ressources choisies.
Classification des résultats	Ordre alphabétique.	Les résultats sont classés par occurrence.
Mise à jour	Aléatoire. En fonction des domaines.	Régulière (pas plus de 100 sites morts).
Navigation	Facile.	Aisée.
Ergonomie	Sobre. Clair.	Convivial. Lecture scan.

CATALOGUES DE BIBLIOTHÈQUES

	<u>Catalogue collectif de France</u> <i>Francophone</i> 	<u>Online Computer Library Center</u> <i>Anglophone</i> 
Auteurs, éditeurs	BNF sous la tutelle des ministères de l'Education Nationale et de la Culture.	Réseau informatisé de bibliothèques fondé en 1967 par les présidents des collèges et universités de l'état d'Ohio (Ohio College Library Center).
Contenu	Informations détaillées sur les bibliothèques françaises, leurs collections et leurs fonds et interrogation de leur catalogue en ligne. Localisation des documents. Moteur de recherche par type de documents et langues.	Possibilité de chercher sur 100 bases de données utilisées par 41 000 bibliothèques dans 82 pays (en France dans plus de 90 bibliothèques universitaires ou publiques). Le site s'adresse à des professionnels de l'information. Moteur de recherches par base.
Classification des résultats	Ordre alphabétique.	En fonction des bases de données.
Mise à jour	Mise à jour très régulièrement et de manière autonome.	Régulière.
Navigation	Facile d'utilisation.	Information dense. Navigation complexe.
Ergonomie	Interface austère.	Aérée et structurée.

BASES DE DONNÉES

C'est un ensemble d'informations organisées autour d'un **même sujet**, directement exploitables et proposées en consultation aux utilisateurs.

	Unité Régionale de Formation et de promotion pour l'Information scientifique et Technique <i>Francophone</i> 	The Internet Archive <i>Anglophone</i> 
Auteurs, éditeurs	Jean-Pierre Lardy : co-directeur de l'URFIST et maître de conférence à l'université de Lyon 1.	Association à but non lucratif Aucune sélection n'est effectuée, le site récupère tous les sites morts sans exception.
Contenu	Une sélection de plus de 800 bases de données. Listes de bases de données. Bibliographies complètes. Consultation gratuite des bases de données mais accès payants aux documents primaires.	Une bibliothèque digitale destinée à conserver tous les documents numériques issus de l'Internet pour les préserver d'une disparition complète. IA fournit des documents créés à partir de 1996 (10 milliards de pages web mais aussi usenet, films et l'ancêtre Arpanet). Dossiers par thèmes.
Classification des résultats	Bases classées par domaines puis par ordre alphabétique.	Classement par occurrence.
Mise à jour	Mise à jour constante des bases de données bibliographiques. Bientôt une nouvelle version du site.	Quotidienne.
Navigation	Simple et structurée.	Informations structurées.
Ergonomie	Claire et agréable.	Couleurs peu attrayantes.

PORTAILS

Ce sont des sites fédérant de nombreuses ressources organisées autour d'un annuaire thématique. Le portail horizontal concerne une activité transversale et le portail vertical (ou vortail) cible une communauté professionnelle et donc un secteur d'activité.

	Enfin (portail horizontal) <i>Francophone</i> 	Objectif Grandes Ecoles (vortail) <i>Francophone</i> 
Auteurs, éditeurs	Auteur : Christophe Hédé, spécialiste des outils de recherche sur Internet. Editeur : IDF www.idf.net/ .	Auteurs : David Chelly et François Lafargue, enseignants à l'Ecole Supérieure de Gestion et à l'Ecole Centrale de Paris.
Contenu	Répertoire les outils de recherche francophones en les classant en grandes catégories. Il est également disponible en version internationale .	Le site Objectif Grandes Ecoles propose une rubrique consacrée aux portails sectoriels.
Classification des résultats	Par occurrence dans les sites sélectionnés	Classés par secteurs d'activité.
Mise à jour	Régulière.	Régulière.
Navigation	Choix de recherche géographique.	Par public et par secteur d'activité.
Ergonomie	Agréable.	Interface très agréable

MOTEURS

Ce sont des systèmes d'interrogation de bases de données. Un **robot** parcourt les documents présents sur le web pour les indexer sur des serveurs.

	Wondir <i>Anglophone</i> 	Turbo10 <i>Anglophone</i> 
Auteurs, éditeurs	Géré par une organisation à but non lucratif composée d'experts en sciences de l'information .	Développé par la société Fleetfoot Internet Solutions Limited basée à Londres. PDG et créateur du moteur : Nigel Hamilton .
Contenu	Associe les possibilités d'un métamoteur et d'un service de recherche humain . Il dispose d'une communauté de volontaires qui répondent aux questions trop complexes pour le moteur de recherche.	Utilise plus de 1000 moteurs de recherches (plutôt anglophones, type Google, Altavista, Lycos...) spécialisés permettant de rechercher dans des bases de données ou des documents du web profond dans des domaines particuliers.
Classification des résultats	- les résultats web - les résultats issus des newsgroups, mailing lists - les propositions de service d'experts de la communauté Wondir - des questions et leurs réponses en relation avec la requête - des dépêches d'actualités liées au domaine de la requête.	Par pertinence.
Mise à jour	Régulière.	Régulière.
Navigation, utilisation	Très simple.	Le module "My Collection", permet de se faire sa propre liste de moteurs figurant dans une liste assez impressionnante (1170 à ce jour).
Ergonomie	Interface conviviale.	Peu attractive.

III/ ÊTRE INCOLLABLE

BIBLIOGRAPHIE

Ouvrages

Recherche et veille sur le web visible et invisible de Béatrice Foenix-Riou, Editions Tec & Doc /Bases Publications avril 2001 234 p. (56.41 €). Cet ouvrage a reçu le prix Idt / Net 2001 Livre Infocom de l'année. Disponible sur le site d' [Amazon](#).

The invisible web finding hidden Internet resources search engines can't see de Chris Sherman et Gary Price, CyberAge Books, septembre 2001, 402 p. (36,67€) www.invisible-web.net/book-info.html

Revues

Netsources : lettre bimestrielle consacrée à la recherche sur Internet. Abonnement d'un an : 120 €.

Bases : lettre mensuelle de référence sur l'actualité du web invisible et des serveurs professionnels. Abonnement d'un an : 230 €.

Ces 2 revues sont éditées par [Fla Consultants](#) ; leur rédactrice en chef est Béatrice Foenix-Riou.

SITES D'EXPERTS

- ✓ [Asktibbs](#) : site d'un documentaliste, Thibaut Deshayes. Nombreux liens vers des sites reconnus traitant de bases de données et d'outils spécialisés ; lettre de diffusion, revue de presse, système de recherche par thème et par mots-clés. Il évalue les sites à l'aide d'étoiles.
- ✓ [Indicateur](#) : les auteurs du site sont Laurent Doublier et Claude Senger ; dossier sur le web invisible. Il propose les rubriques suivantes : qu'est ce que le web invisible ? Moteurs et annuaires de recherche sur le web invisible, les bases de données invisibles et pour en savoir plus.
- ✓ [Berkeley Library](#) : page web à visée pédagogique hébergée sur le site de l'Université de Berkeley (Californie) ; cours sur le web invisible à destination des étudiants de l'Université qui veulent perfectionner leur méthode de recherche sur Internet.
- ✓ Site de [Christophe Asselin](#) (voir 1^{ère} partie du dossier).

NEWSLETTERS

Il est conseillé de s'inscrire aux lettres d'informations suivantes afin de suivre l'actualité de ces outils :

- ✓ [Abondance](#).
- ✓ [Asktibbs](#).
- ✓ Liste de [newsletters](#) recommandées par Christophe Asselin.

Des remarques sur ce dossier, une demande de recherche ou de diffusion personnalisée ?

⇒ Contactez-nous : marion_vacher@yahoo.fr

Dossier mis à jour le 26/05/2004