

Remarque : Vous pouvez consulter les supports de cours (modules B0 à B7) sur le site <http://www.ciimes.org/index.php>. Toutes les évaluations (contrôles continus, QCM et épreuve pratique) se baseront sur ces supports.

Le module correspondant à la séance 4 est le module CIIMES B2 (chapitres 1 à 3) du référentiel C21 (<http://www.ciimes.org/index.php>).

TD4 : Recherche d'information

I. Annuaire

Un annuaire est un outil de recherche qui recense (référence) un très grand nombre de sites en les classant par catégories (ou rubriques) telles que : Actualités et médias, Sports et loisirs etc.... Chaque dernière sous-rubrique est une liste de **sites Web** (et non de pages Web). La construction d'un annuaire est effectuée **manuellement** par la société qui gère l'annuaire. L'utilisateur qui effectue une recherche peut taper un mot clé ou naviguer dans l'arborescence des catégories.

Pour la recherche de mot clé, il faut utiliser :

- des guillemets pour la recherche d'une expression
- le signe + pour la recherche de plusieurs mots
- le signe – pour l'exclusion d'un mot

Il existe également une recherche avancée par mot clé qui permet de préciser quelques paramètres de la recherche

Le plus ancien annuaire www.yahoo.com (américain) est aussi le plus puissant (celui qui répertorie le plus grand nombre de sites. Il existe des versions françaises, espagnoles, allemandes etc... (www.yahoo.fr , www.yahoo.es , www.yahoo.de etc...)

II. Moteur

Un moteur est un outil de recherche qui référence **automatiquement** les **pages Web** se trouvant sur le réseau à l'aide d'un programme appelé spider ou robot. Le contenu de chaque page est analysé et indexé : chaque mot significatif est conservé dans l'index avec l'adresse de sa page et son adresse dans cette page : il est ainsi constitué un index. Les pages sont classées par ordre de pertinence en fonction des mots clés saisis par les utilisateurs du moteur de recherche.

L'utilisateur qui effectue une recherche présente une requête, celle-ci est une combinaison de mots-clés et d'opérateurs logiques. Voici quelques règles pour le moteur de recherche Google, qui est actuellement le moteur le plus performant :

- l'opérateur AND est ajouté implicitement
- les termes ignorés sont « le, la, les, du, avec, vous, etc »
- le signe + correspond à l'opérateur AND
- le signe * indique que l'utilisateur souhaite avoir n'importe quelle lettres ou chiffres à la place de *
- il n'y a pas de différence avec/sans accent, avec/sans ç, majuscule/minuscule
- le signe – permet d'exclure un mot
- les guillemets permet de rechercher des expression

- pour limiter la recherche à un site ou à un domaine spécifique, il faut utiliser la syntaxe « site:domaine.com » dans le champ de recherche de Google après l'expression cherchée
- il existe différents types de recherche possible avec les onglets: web, image, actualités, groupe....

Il existe également une recherche avancée qui permet de limiter la recherche a certaines pages choisies, de préciser le champ de recherche, etc...

III. Annuaire ou Moteur

Annuaire: recherche le mot de la requête donnée parmi les mots clés rattachés aux sites qu'il référence avec un classement des sites par thèmes

=> donne moins de résultats qu'un moteur, mais ces résultats seront plus pertinents.

Moteur de recherche = recherche le mot de la requête donnée dans toutes les pages qu'il référence automatiquement

=> donne nettement plus de résultats mais beaucoup ne seront pas pertinents.

IV. Catalogue

Au sens général, un catalogue est une **liste**. Par exemple, pour une bibliothèque, le catalogue correspond à la liste de tous les documents de la bibliothèque.

Il faut savoir qu'il existe un catalogue collectif des Universités françaises appelé le **SUDOC** (Système Universitaire de Documentation : <http://corail.sudoc.abes.fr/>) et qui regroupe plus de 5 millions de références.

Le but d'un catalogue est donc de permettre d'identifier un document (accès à la notice bibliographique : fiche descriptive du document) puis de le localiser (accès à la liste des bibliothèques possédant le document et la cote de celui-ci).

Un catalogue peut contenir des **indexes**. Un index est une liste élaborée à partir des éléments de description des documents. On aura ainsi des index alphabétiques des auteurs, des titres, des sujets et des index numériques des ISBN (International Standard Book Number : numéro international unique attribué à un livre) et des ISSN (International Standard Serial Number : numéro international unique attribué à un titre de revue)... L' ISSN est le moyen le plus efficace pour localiser une revue dans un catalogue.

Il est possible de faire une recherche simple en choisissant un index puis en saisissant un ou plusieurs mots, ou une expression entre guillemets. Une recherche avancée est également possible sur une combinaison d'index (ex : collection et mots du sujet)

V. Récupérer des images sur le Web

Pour enregistrer une image d'une page Web affichée dans un dossier, il faut cliquer avec le bouton droit de la souris sur l'image à sauvegarder puis dans le menu contextuel qui apparaît, cliquez sur la commande « Enregistrer l'image sous... ».

VI. Récupérer une page HTML

Pour récupérer une page Html, vous pouvez sauvegarder, sous la forme d'un fichier sur le support de votre choix, la page web que vous avez affichée (menu "Fichier/ Enregistrer la page sous..."). Vous avez le choix d'enregistrer la page web complète texte + images, icônes (ainsi que le fichier de style lié à la page) ou la page html uniquement. Sélectionnez l'option dans "Type".

VII. Télécharger un fichier ou un logiciel

Il est également possible de télécharger un fichier ou un logiciel. Différents protocoles d'Internet permettent de télécharger des fichiers images, sons, fichiers au format PDF, fichiers compressés, etc. En général il s'agit d'un simple lien vers le fichier.

Pour télécharger des logiciels, vous pouvez utiliser les sites
« <http://www.01net.com/telecharger/> » et « <http://www.download.com/> ».

VIII. Lire un fichier multimédia enregistré sur le Web

Pour lire des fichiers multimédias enregistrés sur le web, il faut utiliser des logiciels spécifiques aux formats des fichiers.

Voici quelques formats de fichiers son :

- **MP3** : format de musique compressée, très utilisé
- **AIF, .AIFF, .AIFC** : Les fichiers portant l'extension .aif, .aiff, .aifc sont des clips sonores, ils peuvent être lus grâce à un lecteur multimédia.
- **MOV** : Fichier QuickTime Movie
- **RAM, .RA** : Fichier au format Real Audio
- **WMA** : Format de musique compressée, format propriétaire de la société Microsoft.. Les fichiers ont une moins bonne qualité que le MP3. Certains peuvent être protégés par une licence
- **MID/MIDI** : fichiers uniquement de musique générés par des instruments reliés à la carte son du PC
- **WAV** : Format de musique très répandu, fichiers non compressés.

Il existe plusieurs logiciels lisant de nombreux formats : « Winamp », « iTunes », « Windows Media Player »...

Pour des fichiers vidéo, il faut un équipement en logiciel de lecture, généralement commun aux fichier audio, comme par exemple « Windows Media Player », « Quicktime », « RealPlayer », « BSPlayer »... Deux modes de lecture permettent de lire différents types de fichiers :

- le streaming, qui traduit la notion de flux vidéo accessible dès la connexion et alimenté au fur et à mesure de la diffusion : une connexion au réseau haut débit (ADSL, Câble ...) est nécessaire dans ce cas pour permettre un transfert des données suffisamment rapide pour conserver la continuité du flux de lecture.
- le téléchargement du fichier suivi de la lecture en différé.

Il existe différents formats pour les fichiers vidéo (AVI, MOV, MPEG, WMV, DIVX, XDIV,...). Pour obtenir ces formats de fichiers, on procède à une compression des images en utilisant des algorithmes permettant de réduire significativement les flux de données en compressant/décompressant les données vidéos. On appelle ces algorithmes **CoDec** (pour COmpression/DECompression).

IX. Aspirer un site

Aspirer un site c'est copier tous les fichiers d'un site Web. Cela peut se faire avec un logiciel qui enregistre les fichiers dans un répertoire donné comme Memoweb, eCatch, WebZip, HTTrack, La structure complète du site est ainsi accessible hors connexion sur l'ordinateur, mais on perd toutes les informations dynamiques, notamment celles issues d'une base de données. En général, il suffit d'entrer l'adresse du site à aspirer et de suivre les indications de l'interface (détermination de la profondeur à laquelle vous irez chercher les informations dans un site donné,...).

Attention :

Il faut respecter les droits de propriété intellectuelle lors de l'utilisation de sites aspirés.

X. Exercice

Allez sur le site du journal « Le Monde » ([url : www.lemonde.fr](http://www.lemonde.fr)) et choisissez un article. Puis récupérez toutes les informations que vous trouverez sur ce site à propos de l'article choisi (commentaires, images, photos, texte...). Intégrez alors ses informations à un document Word structuré.

Allez chercher d'autres informations de différents types de formats (photos, dessins, textes, articles, livres...) sur l'article choisi ou autour du sujet de l'article en utilisant différents moteurs de recherche et annuaires. Récupérez également les résultats pertinents (avec quelques informations sur chaque document comme l'auteur, le lieu où l'on peut le trouver...) obtenus grâce à un catalogue comme le SUDOC. Puis insérez les informations trouvées au document Word précédemment créé en les classant (par type de formats par exemple ou par type de sources...). Pensez à indiquer les sources utilisées.

Vous trouverez après le glossaire des exemples d'adresses de moteurs et d'annuaires de recherche.

Glossaire

URL : (Uniform Ressource Locator) Adresse unique qui permet d'accéder à un site ou une page web à partir de sa saisie dans la barre d'adresses du navigateur.

Navigateur ou **logiciel de navigation** : Dans un environnement de type Internet, logiciel qui permet à l'utilisateur de rechercher et de consulter des documents, et d'exploiter les liens hypertextes qu'ils comportent. Les navigateurs connus sont "Internet Explorer" et "NetScape".

Lien hypertexte : Mot ou ensemble de mots qui par un simple clic permet sur une page Web ou un e-mail d'accéder à une autre partie de la page, à une autre page ou à un autre site. Une image cliquable peut également servir de lien.

Moteurs de recherche et annuaires Français

Moteurs de recherche	Altavista France : http://fr.altavista.com	Google France : http://www.google.fr
	Hotbot France : http://www.hotbot.fr	Alltheweb : http://www.alltheweb.com
	Free : http://www.free.fr	Abacho : http://www.abacho.fr
	Reacteur.com : http://www.reacteur.com	
Annuaires et moteurs de recherche	Voila : http://www.voila.fr	Yahoo France : http://fr.yahoo.com
	Lycos France : http://www.lycos.fr	Aol France : http://www.aol.fr
	Nomade Monde : http://www.nomade.tiscali.fr	

Moteurs de recherche et annuaires Anglais : (Angleterre et USA)

Moteurs de recherche	Altavista.com : http://www.altavista.com	Altavista.co.uk : http://uk.altavista.com
	Google.co.uk : http://www.google.co.uk	Alltheweb : http://www.alltheweb.com
	Netscape.com : http://www.netscape.com	Hotbot.com : http://www.hotbot.com
	Msnsearch.com : http://www.msn.com	Lycos.com : http://www.lycos.com
	DirectHit : http://www.directhit.com	Euroseek.com : http://www.euroseek.com
	Goto.com : http://www.goto.com	Wisenut : http://www.wisenut.com
	Aesop : http://www.aesop.com	Temoa : http://www.teoma.com
	Dmoz : http://www.dmoz.com	Yahoo.com : http://www.yahoo.com

Outil de référence en économie, comptabilité et gestion disponible à la Bibliothèque Universitaire de Lyon2 (<http://scd.univ-lyon2.fr>):

Econlit : Cédérom de références bibliographiques en économie depuis 1969 ; il comprend les notices réalisées par le Journal of Economic literature.