

# Chapitre 5

---

## *Recherche, utilisation et récupération de l'information*

---

<b>5.1. Les annuaires de recherche</b> .....	1
5.1.1. Etude de l'annuaire YAHOO .....	2
5.1.2. Etude du moteur GOOGLE .....	2
<b>5.2. Les moteurs de recherche</b> .....	3
<b>5.3. Le langage d'interrogation et de recherche</b> .....	3
<b>5.4. Récupération et impression d'une page HTML</b> .....	5
<b>5.5. Récupérer une image</b> .....	7
<b>5.6. Télécharger un fichier ou un logiciel</b> .....	8
<b>5.7. Lecture d'un fichier HTML en local</b> .....	9
<b>5.8. Lecture d'un fichier multimédia enregistré sur le Web</b> .....	9

---

## Chapitre 5 : Recherche, utilisation et récupération de l'information

### I. Les annuaires de recherche :

- Un annuaire (on parle parfois de guide) est un outil de recherche qui recense un certain nombre de sites.
- Un annuaire est un sous-ensemble du WEB.
- L'annuaire est constitué de rubriques (appelées aussi catégories) telles que : Actualités et médias, Sports et loisirs, etc.
- Chaque rubrique est subdivisée en sous-rubriques plus précises, elles-mêmes subdivisées en sous-sous-rubriques, etc.
- Chaque dernière sous-rubrique est une liste de sites Web (et non de pages Web).

Deux principaux modes :

- En naviguant (browsing) de rubrique en sous-rubrique, jusqu'aux sites accrochés dans l'annuaire : on est ainsi guidé vers une liste de sites centrés sur le thème de recherche ;
- Par mots-clés (searching) : ces mots sont entrés dans une boîte de recherche au niveau d'une rubrique ou sous-rubrique de l'annuaire : on obtient comme résultat des rubriques de l'annuaire et des sites. L'annuaire effectue sa recherche sur les noms de rubriques de l'annuaire, les noms des sites et le bref descriptif (non visible pour l'utilisateur) associé à chaque site.

#### Avantage :

- La navigation est très simple et constitue un guidage très efficace ;
- Les sites ayant été sélectionnés, dans les 2 modes de recherche les sites trouvés sont de bonne qualité et bien centrés sur le thème cherché;

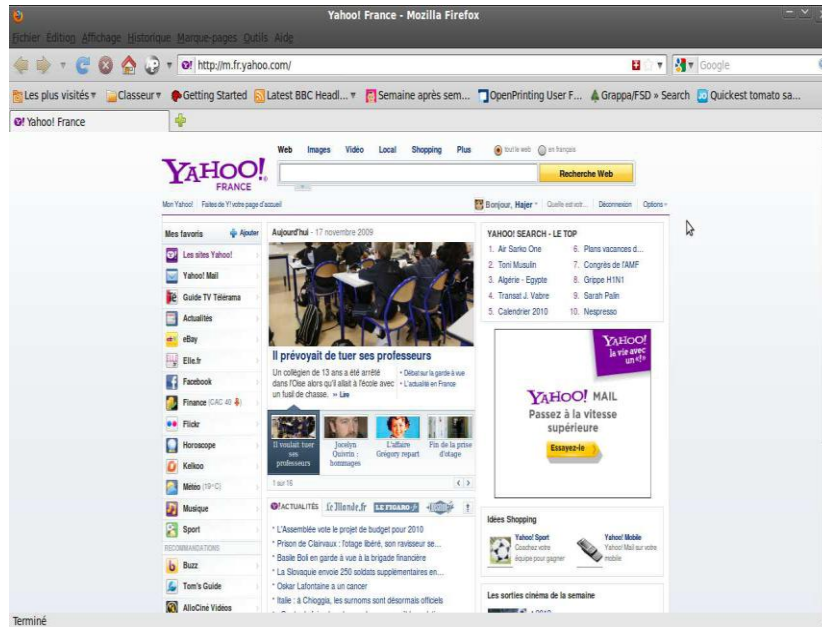
#### Inconvénients :

- La taille de l'annuaire est réduite et ne suit pas d'assez près la croissance du Web ;
- La tenue à jour de l'annuaire n'est pas très bonne (nouveaux sites, sites disparus, etc.)

## I.1. Etude de l'annuaire YAHOO :

Le plus ancien annuaire

- La recherche par navigation (browsing)
- La recherche simple par mots-clés (searching)
- La recherche avancée par mots-clés (searching)



## I.2. Etude du moteur GOOGLE :



Sa page d'accueil est d'une extrême simplicité ;

- Il est très puissant : en 2006, Google a dans son index plus de 8 Milliards de pages ;
- Il est très rapide : le temps de réponse moyen à une requête est de 0,5 seconde ;
- Le classement de ses résultats est très efficace : les meilleurs résultats de ses recherches sont placés en tête.

## II. Les moteurs de recherche :

Le moteur de recherche fonctionne sur un système radicalement différent de celui de l'annuaire, qui est celui de l'indexation:

- Des robots logiciels (appelés crawlers ou spiders) parcourent le Web, vont de page en page (en fait de lien en lien) et au fur et à mesure de leur parcours, indexent le contenu textuel des pages (il s'agit ici de pages Web et non de sites) rencontrées. 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.
- Un moteur de recherche est un logiciel qui a construit l'index de la totalité (théoriquement) des pages du Web : cette indexation est totalement automatique.

### Fonctionnement :

- L'utilisateur, pour faire une recherche, présente une requête, c'est-à-dire une combinaison de mots-clés et d'opérateurs logiques tels que : ET, OU, SAUF etc.
- Le moteur agit en 2 étapes :
  - ✓ Sélection des pages satisfaisant à la requête ;
  - ✓ Présentation des adresses des pages après classement par le moteur par pertinence décroissante : Le moteur a puisé les URLs de ces pages dans son index.

### Avantage :

- La taille de l'index augmente avec la taille du WEB, à cause de l'efficacité du robot d'indexation ;
- La mise à jour de l'index est suffisamment fréquente ;
- Le nombre de résultats en réponse à une requête est souvent énorme.

### Inconvénients :

- La réelle difficulté de choisir les bons mots-clés, en particulier lorsque la recherche est faite vers des pages en langue étrangère ;
- La sélection des pages correspondant à une requête étant automatique, certaines de ces pages seront non pertinentes par rapport au thème de recherche.

## III. Le langage d'interrogation et de recherche :

Les outils de recherche ne comprennent pas le langage naturel.

Formaliser une requête de recherche consiste à transcrire une question dans un langage compris par les outils de recherche.

Ce langage comporte :

- ✓ Des mots clés
- ✓ Et l'organisation de ceux ci dans une syntaxe adaptée à l'outil de recherche que vous utilisez.

### **III. 1. De l'importance des mots clés :**

Les concepts déterminés lors de la préparation de la recherche sont traduits en **mots clés**.

Le mot clé condense l'idée que vous voulez exprimer. Vous pouvez donc avoir besoin de plusieurs mots clés pour résumer ou préciser une idée.

Sur Internet, le mot clé est une chaîne de caractères issue du langage naturel.

Lorsque l'outil effectue une recherche, il compare la chaîne de caractères demandée avec les chaînes de caractères stockées dans son index. Le sens des mots est donc inconnu des outils de recherche.

### **III. 2. Les équations de recherche ou comment écrire sa question sur les outils de recherche :**

Pour combiner c'est à dire associer plusieurs mots clés dans une recherche, vous disposez de ce que l'on appelle des **opérateurs de recherche**. Pour l'outil que vous utilisez, ces opérateurs sont comme des codes qui lui indiquent ce que vous souhaitez faire.

### **III.3. Les opérateurs booléens : « ET », « OU », « SAUF » :**

Le terme de « booléen » vient du mathématicien George Boole qui a travaillé sur la théorie des ensembles. La recherche dans les index des outils est fondée sur la théorie des ensembles.

Un ensemble est le résultat d'une requête.

L'intersection marquée par le « **ET** » permet de trouver les termes communs aux ensembles visés.

- ✓ La réunion marquée par le « **OU** » permet de trouver les mots clés recherchés appartenant aux deux ensembles visés, ou à l'un ou à l'autre.
- ✓ L'exclusion marquée par le « **SAUF** » qui permet d'interdire un mot clé et d'éliminer tous les documents que l'on ne souhaite pas obtenir.

### **III.4. Les opérateurs de proximité :**

**La recherche de proximité exacte** peut être demandée en utilisant des guillemets. Ces derniers encadrent une expression ou une suite de mots clés que l'on souhaite trouver ainsi, écrits dans cet ordre.

### **III.5. La troncature :**

Marquée par le signe « **\*** » placé à la fin d'un mot, elle permet d'élargir la recherche à tous les mots qui ont une racine commune.

### **III.6. Les filtres de recherche :**

Les outils de recherche permettent aussi de localiser les informations recherchées selon différents paramètres.

La recherche des mots-clefs peut se faire dans :

- ✓ le titre d'une page
- ✓ l'adresse d'une page
- ✓ le nom de domaine
- ✓ le format des documents
- ✓ les liens hypertextes...

Tous ces champs sont interrogeables. Ces filtres sont souvent proposés dans les parties « **recherche avancée** » des outils de recherche.

### III. 7. La méthode d'interrogation sur Internet :

La recherche sur Internet est intuitive.

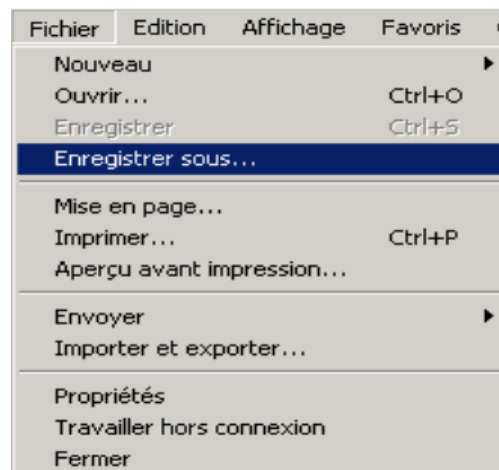
Il est conseillé de travailler **par étapes** plutôt que d'écrire de longues équations avec des « ET » et des « OU » accompagnés de parenthèses. Si cela est techniquement possible, en revanche, les résultats sont souvent nuls ou peu pertinents sans que l'on sache toujours pourquoi.

Il faut savoir s'adapter aux possibilités de navigation, à l'existence de liens entre les informations et aux structures mêmes des contenus des pages que vous allez visiter. Il faut être prêt à réagir, quitte à réajuster sa recherche, en fonction des premiers résultats obtenus.

### IV. Récupération et impression d'une page HTML :

Vous pouvez sauvegarder sur votre poste local la page web que vous avez affichée, sous la forme d'un ou plusieurs fichiers.

La plupart des navigateurs vous proposent l'option "Enregistrer la page sous...". Déroule le menu "Fichier" puis cliquez sur "Enregistrer sous...".

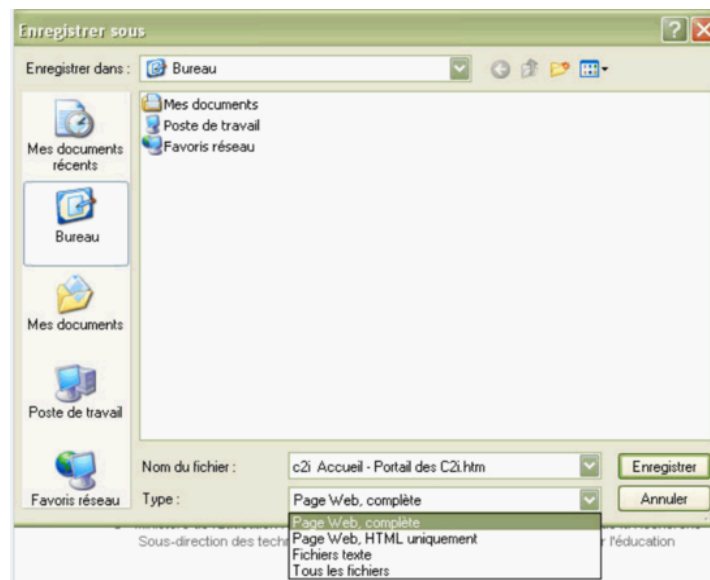


Dans la fenêtre "Enregistrer sous" sélectionnez l'emplacement où vous souhaitez sauvegarder la page.

Repérez le nom de fichier sous lequel vous pourrez par la suite retrouver votre sauvegarde puis cliquez sur le bouton "Enregistrer". Vous avez le choix entre :

- enregistrer la page web complète (texte, images, icônes, ainsi que le fichier de style lié à la page)
- enregistrer la page html uniquement.

Pour faire ce choix, sélectionnez l'option retenue dans « **Type** ».



Si vous souhaitez par la suite consulter cette page, rendez vous dans le répertoire sélectionné et double-cliquez sur le nom du fichier sous lequel vous l'avez enregistrée.

### **Remarque :**

Les sites qui proposent du contenu interactif ne peuvent être visualisés correctement que si la connexion à Internet est établie. Hors connexion, certains éléments de la page ne seront pas visibles.

### **Imprimer une page HTML :**

#### **Méthode :**

Lorsque vous désirez imprimer une page Web, déroulez le menu "Fichier" du navigateur puis cliquez sur "Imprimer".

La fenêtre "Impression" s'ouvre, sélectionnez votre imprimante. Vous imprimerez la totalité de la page web, vous pouvez également définir la zone d'impression et le nombre de copies.

**Remarque :**

On peut rencontrer quelques anomalies lors de l'impression d'une page web.

Pour une impression correcte, certains sites proposent une mise en page spécifique, accessible dans la page via un lien « **Imprimer** » ou une icône représentant une imprimante.

**V. Récupérer une image :**

On peut également choisir d'enregistrer non pas une page html entière mais simplement une image contenue dans cette page.

Cliquez alors avec le bouton droit de votre souris sur l'image que vous désirez sauvegarder.

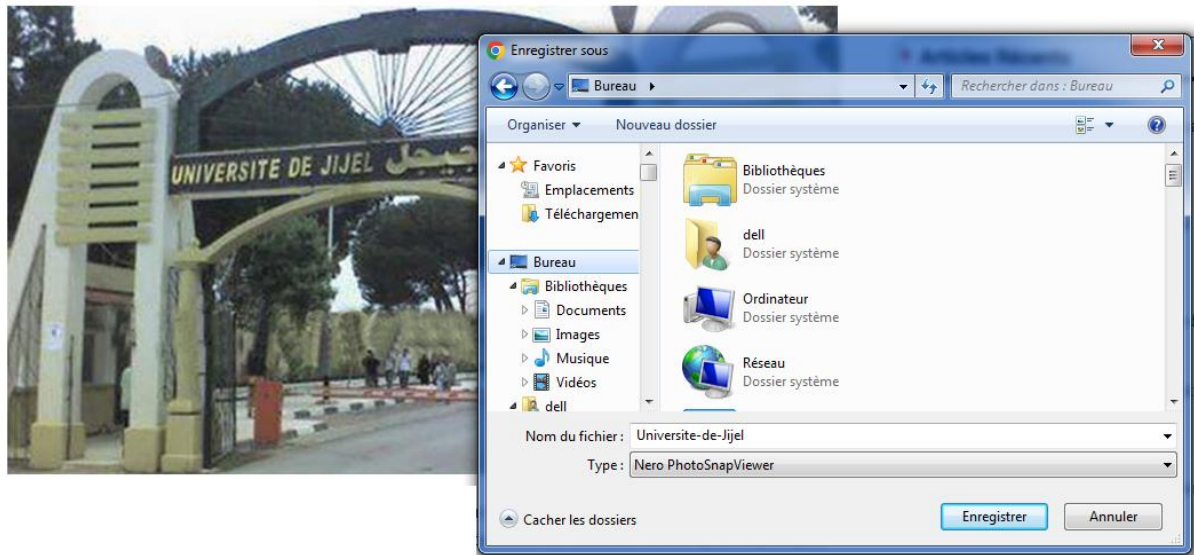


Dans le menu contextuel qui apparaît, cliquez sur la commande « Enregistrer l'image sous ».

Sélectionnez le répertoire de stockage choisi dans l'arborescence.

- Précisez un nom pour l'image dans « **Nom du fichier** ».
- Choisissez le « **Type** » à l'aide du menu déroulant.

- Cliquez sur « **Enregistrer** ».



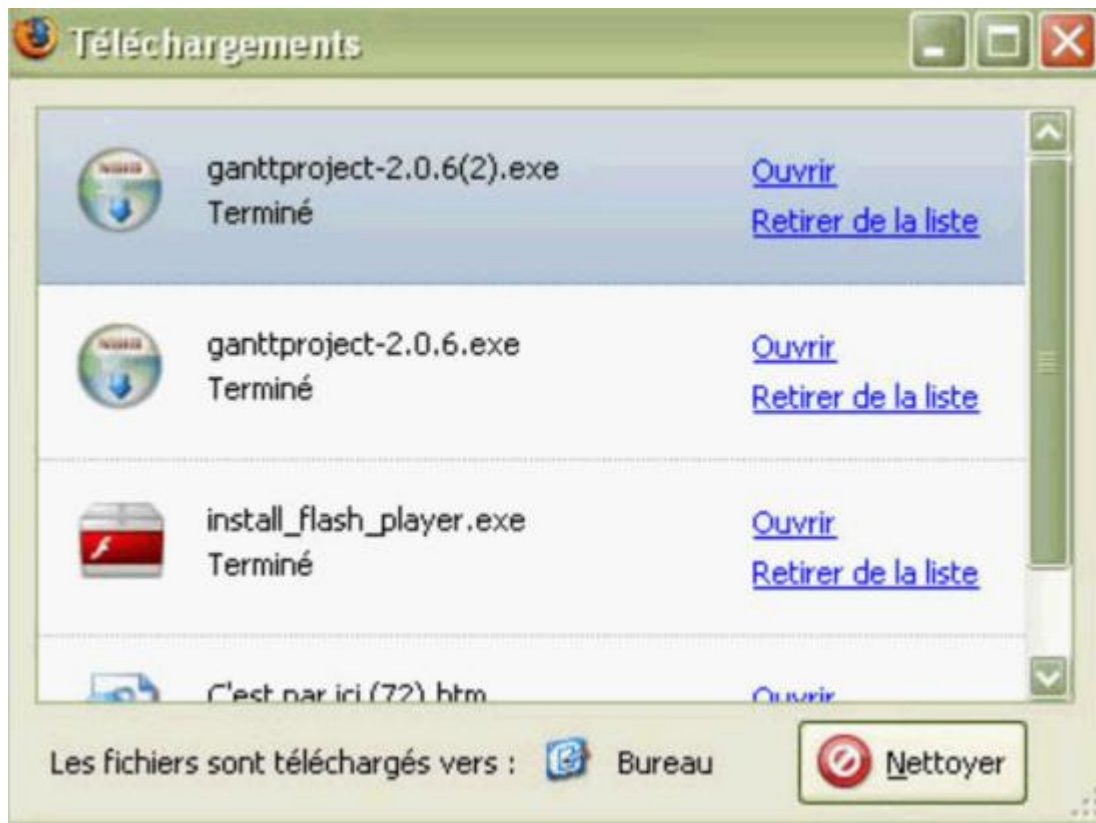
## VI. Télécharger un fichier ou un logiciel :

Différents protocoles d'Internet permettent de télécharger des fichiers images, sons, PDF, fichiers compressés, ou des logiciels.

Un simple lien vers le fichier ou l'application à télécharger permet de lancer le téléchargement du fichier ou de l'application.



Choisissez « **Enregistrer le fichier** », celui-ci étant enregistré, par défaut dans le répertoire « **Téléchargements** », puis cliquez sur « **Ouvrir** ».



### VII. Lecture d'un fichier HTML en local :

Vous pouvez ouvrir un fichier html préalablement enregistré en le chargeant dans un navigateur Internet :

- Soit en utilisant le menu « **Fichier** » puis la commande « **Ouvrir** ».
- Soit en faisant un double clic sur le fichier html.

### VIII. Lecture d'un fichier multimédia enregistré sur le Web :

#### ❖ Formats de fichiers son :

Il existe différents types de formats de fichiers audio :

- MP3 : format de musique compressée, très utilisé
- .aif, .aiff, .aifc : extension désignant des clips sonores (nécessite de ce fait un lecteur multimédia).
- .mov : fichier QuickTime Movie
- ram, .ra : fichier au format Real Audio

- WMA : format de musique compressée, appartenant à Microsoft. Les fichiers ont une moins bonne qualité que le MP3. Certains peuvent être protégés par une licence.
- MID/MIDI : fichiers comportant des instructions pour jouer des notes de musique (en quelque sorte, il s'agit d'une partition).
- WAV : format de musique très répandu, fichiers non compressés.

Cette liste n'est pas exhaustive.

### ❖ Lire des fichiers son

Il existe plusieurs logiciels téléchargeables gratuitement, permettant de lire les fichiers audio :

- Winamp semble le plus utilisé. Il est disponible dans les environnements Windows et Mac, et téléchargeable à l'URL suivante : <http://www.winamp.com/>
- Windows Media Player est le logiciel de Windows qui ne devrait plus à l'avenir être préinstallé. Il peut être téléchargé à l'adresse suivante : [http://www.01net.com/windows/Multimedia/lecteurs\\_video\\_dvd/fiches/24011.html](http://www.01net.com/windows/Multimedia/lecteurs_video_dvd/fiches/24011.html)
- I-tunes, le lecteur audio d'Apple rendu célèbre grâce à sa gestion des I-pod, est également très utilisé : <http://www.apple.com/fr/itunes/download/>

### ❖ Formats de fichiers vidéo

La vidéo numérique consiste à afficher une succession d'images numériques à une cadence variable par unité de temps (25 images par seconde pour la plus haute qualité).

On procède à une compression des images basée :

- sur la perception des informations visuelles de façon subjective et en prenant en compte leur interprétation, par notre cerveau, selon des références culturelles,
- sur des algorithmes permettant de réduire significativement les flux de données en compressant / décompressant les données vidéos notamment selon le principe de **redondance temporelle**.

On appelle ces algorithmes CoDec (pour COmpression/ DECompression).

Parmi les différents systèmes existants, MJPEG et MPEG sont normalisés par l'ISO. Cet organisme de normalisation a institué deux groupes de travail :

- Le JPEG (Joint Photographic Expert Group) créé au départ pour les images fixes.
- Le MPEG (Moving Picture Expert Group) pour les vidéos.

### ❖ Lire des fichiers vidéo

Lire de la vidéo sur le web suppose un équipement en logiciel de lecture, généralement commun aux fichiers audio, et une connexion au réseau haut débit (ADSL, câble...) permettant un transfert des données suffisamment rapide pour conserver la continuité du flux de lecture.

Deux modes de lecture permettent de lire différents standards :

- le streaming qui traduit la notion de flux vidéo accessible dès la connexion et alimenté au fur et à mesure de la diffusion,
- le téléchargement du fichier suivi de la lecture en différé.

Vous trouverez ci-dessous une liste de lecteurs vidéo téléchargeables gratuitement.

BSplayer	Lecteur multimédia aux fonctions de base compatible avec de nombreux formats (y compris le format DivX)
Windows Media Player	Lecteur issu de l'environnement Microsoft.
Quick Time	Lecteur multimédia également très utilisé, issu de l'environnement Apple
Real Player	Pionnier de la Vidéo en streaming.
VLC	Lecteur multimédia de référence, léger et portable permettant de lire de formats vidéo, audio et streaming.
Media Classic Player	Lecteur aux fonctions de basse.