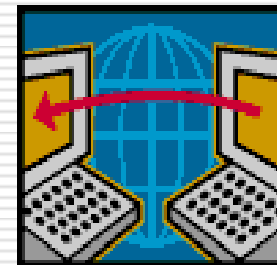
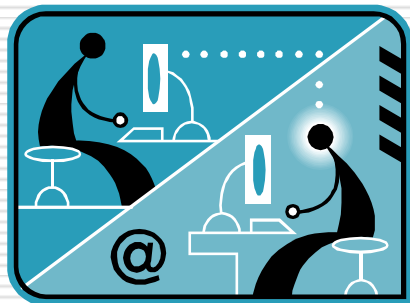


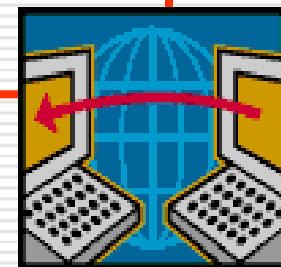
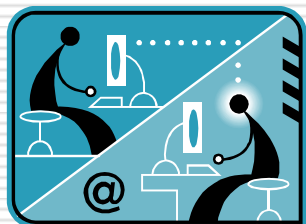


C2i
**Certificat Informatique
et Internet**
**D4 : recherche d'information
à l'ère du numérique**





- **1/ Internet, le web**
- **2/ Les outils de recherche**
 - Moteurs, annuaires, répertoire de signets, portails
- **3/ Analyser la pertinence des résultats**
- **4/ Méthodologie de recherche**
- **5/ Référencer et récupérer les ressources**



❑ On appelle «**Web**» (nom anglais signifiant «**toile**»), contraction de «World Wide Web» (d'où l'acronyme *www*), la possibilité offertes par le réseau Internet de naviguer entre des documents reliés par des liens hypertextes, notamment grâce au protocole HTTP.

Pour naviguer entre des documents (appelés «**pages web**») on utilise un logiciel appelé **navigateur**.

❑ 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 :

- **Le nom du protocole** : c'est-à-dire en quelque sorte le langage utilisé pour communiquer sur le réseau. Le protocole le plus largement utilisé est le protocole HTTP (HyperText Transfer Protocol),
- **Le nom du serveur** : nom de domaine de l'ordinateur hébergeant la ressource demandée.
- **Le chemin d'accès à la ressource** : Cette dernière partie permet au serveur de connaître l'emplacement auquel la ressource est située, c'est-à-dire de manière générale l'emplacement (répertoire) et le nom du fichier demandé

<http://portail-formation.cnam.fr/catalogues-de-la-formation/>

Parmi les sources d'information sur le web, on distingue :

❑ **les sources d'information traditionnelles** (agences de presse, grandes organisations d'intérêt privé ou public, associations d'experts, publications scientifiques, etc.) qui diffusent des informations validées selon un processus reconnu ;

❑ L'agence de presse AFP, l'organisation internationale OCDE, la Bibliothèque Nationale de France, les éditeurs de publications scientifiques sont des sources d'information traditionnelles.

❑ **les sources émergeant de structures collectives plus ou moins formelles**, dont les règles de publication et de validation de l'information sont plus ou moins strictes.

❑ L'encyclopédie en ligne Wikipédia est une source qui émerge d'une communauté d'auteurs ; la qualité de l'information qu'elle diffuse est liée aux règles et processus d'édition qui lui sont propres.

❑ **les sources informelles** (pages web personnelles, blogs d'internautes, etc.) publiées sans aucun contrôle.

❑ Le blog d'un prix Nobel d'économie relève des sources informelles. Les informations qui y sont publiées peuvent être d'une grande valeur, sans toutefois bénéficier d'une validation par d'autres spécialistes du domaine.

Mise en pratique : rechercher des informations sur la biographie de François Rabelais. Identifier les différents types de sources.

Les bibliothèques conservent des documents. Pour les répertorier et les localiser, ils sont décrits par leur notice bibliographique.

Une **notice bibliographique** est la description d'un document. Elle est organisée selon des **champs** : auteur, titre, éditeur, collection, ISBN, année de publication, type de publication, résumé, mots-sujet, etc.

Un **catalogue de bibliothèque** donne accès aux notices bibliographiques des documents qu'elle contient.

Pour permettre une recherche par le sujet, les documents doivent être **indexés**, c'est-à-dire associés à des **mots-clé**.

☐ **indexation manuelle** s'appuyant sur un **thésaurus** ; c'est un ensemble de mots-clé ou mots-sujet prédéfinis que le bibliothécaire choisit d'associer au document pour décrire son contenu.

☐ **indexation automatique** des divers champs des notices, notamment du champ « Titre » et du champ « Résumé ».

Le catalogue du **SUDOC** (Système Universitaire de Documentation) est le catalogue collectif français réalisé par les bibliothèques et centres de documentation de l'enseignement supérieur et de la recherche.

Un **annuaire de recherche** est un outil de recherche qui répertorie des sites web sélectionnés pour leur qualité par des spécialistes du domaine. Les sites sont organisés dans une hiérarchie de **catégories**.

Le processus d'alimentation d'un annuaire reposant sur une **sélection humaine**, les ressources qu'on y trouve sont théoriquement de bonne qualité.

❑ les **annuaires généralistes**

Open Directory Project – DMOZ

❑ les **annuaires spécialisés** : :

Signets de la BNF (sélection des bibliothécaires de la Bibliothèque Nationale de France)

Infomine (ressources académiques)

Application : explorer DMOZ avec les élèves

Un **moteur de recherche** est un outil permettant de retrouver des pages web à partir d'une requête.

C'est un logiciel qui réalise les tâches suivantes :

- ❑ **moissonnage** : les pages du web sont parcourues automatiquement par un robot ;
- ❑ **indexation automatique** : les pages sont analysées pour en extraire les mots significatifs ;
- ❑ **interrogation** en fonction d'une requête : les mots clés de la requête sont comparés aux mots extraits par l'indexation et une liste de pages web sélectionnées est affichée par ordre de pertinence.

Google, Yahoo Search ou Bing sont des moteurs de recherche généralistes.

Il existe des **moteurs de recherche spécialisés** dans certains types de ressources (images, vidéos, publications académiques, blogs, flux RSS, personnes, etc.) ou dans certains domaines (technologie, musique, mathématiques, santé, chimie, etc.).

Exemple



Indexation d'écrits universitaires

❑ *Pour rechercher une information via un moteur de recherche ou un catalogue de bibliothèques, on écrit une requête composée de mots-clés liés par des opérateurs booléens.*

Opérateurs booléens	Inclusion : <ul style="list-style-type: none">▪ AND (opérateur implicite) ou utiliser le signe (+) Union : <ul style="list-style-type: none">▪ utiliser la syntaxe(OR), l'opérateur doit être saisi obligatoirement en majuscules Exclusion : <ul style="list-style-type: none">▪ utiliser le signe (-)
Recherche d'une expression exacte	Utiliser les guillemets « »

- ❑ *lapin + lievre : les pages contenant les mots lapin et lièvre*
- ❑ *lapin OR lievre : les pages contenant les mots lapin et lièvre, lapin seul, lièvre seul*
- ❑ *lapin – lievre : les pages avec lapin seul*
- ❑ *« lapin et lievre » : exactement cette expression*
- ❑ *lap* : toutes les pages contenant lap suivi de quelque chose*

Application : formuler des requêtes dans un moteur de recherche et un catalogue de bibliothèques.



Recherche avancée

Trouvez des pages
avec...

tous les mots suivants :

Pour effectuer cette opération dans le champ de recherche

Saisissez les mots importants : terrier tricolore

ce mot ou cette expression
exact(e) :

Ajoutez des guillemets autour des mots exacts : "terrier"

l'un des mots suivants :

Saisissez OR entre tous les mots à inclure : miniature OR standard

aucun des mots suivants :

Placez un signe - (moins) devant les mots à exclure :
-rongeur, -"Jack Russell"

nombres compris entre :

et

Placez deux points entre les nombres, et ajoutez une unité de mesure :
10..35 kilos, 300..500 USD, 2010..2011



Les signets

Quelques **répertoires de signets** alimentés par des professionnels :

> www.signets-universites.fr

Les principales ressources du Web validées pour leur légitimité scientifique, classées par grands domaines puis sous-domaines

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

Une sélection commentée de ressources accessibles par Internet, choisies par les bibliothécaires de la Bibliothèque Nationale de France.

> http://www.netvibes.com/kiosquelycee#La_Une

Netvibes sur la presse destiné aux lycéens et administré par un professeur documentaliste de l'Académie de Versailles.

La veille informationnelle consiste à se tenir au courant de l'actualité d'un domaine ou d'un sujet.

Pour faciliter ce travail, on peut :

- ☐ Visiter soi même les sites qui nous intéressent, mémorisés dans les signets : Méthode PULL (on tire l'information vers soi)

- ☐ Utiliser des outils qui vont automatiquement chercher les nouveautés sur un sujet à votre place : Méthode PUSH (l'information est poussée vers vous).



- ☐ agréger les flux d'actualités diffusés par différents sites ;

Cela permet d'afficher les derniers articles dans un même espace pour éviter de visiter les différents sites un par un.

- ☐ s'abonner à des lettres d'information ou *newsletters* ;

Une **lettre d'information** électronique, ou *newsletter*, diffuse régulièrement par courriel des informations sur un sujet donné.



- ❑ agréger les flux d'actualités diffusés par différents sites ;

Cela permet d'afficher les derniers articles dans un même espace pour éviter de visiter les différents sites un par un.

- ❑ s'abonner à des lettres d'information ou *newsletters* ;

Un **flux d'informations** est une technologie permettant de publier une succession d'articles ou de contenus multimédia diffusés par un site web sur une autre URL du web. On parle aussi de **fil d'actualité** ou de fil d'information.

Le terme **flux RSS** (Real Simple Syndication) désigne la technologie la plus répandue de flux web.

La lecture d'un flux RSS se fait par un **agrégateur de flux**. C'est un logiciel qui se connecte sur les sites d'origine et permet d'afficher dans un même espace toutes les actualités en provenance de flux d'informations du web.



Les navigateurs intègrent un agrégateur

Logiciels spécialisés : exemple google alerte <http://www.google.fr/alerts?hl=fr>

Site lpo.fr : abonnement flux et lettres d'informations

-
- ❑ *Traces mémorisées sur l'ordinateur lors de la navigation sur le web :*
 - ❑ la liste des URL consultées (**historique**) ou sauvegardées (**signets, favoris** ou **marque-pages**) ;
 - ❑ les fichiers que le navigateur télécharge pour l'affichage des pages web (**cache du navigateur**) ;
 - ❑ les fichiers que l'internaute a téléchargés (**téléchargement**)
 - ❑ Les mots de passe

 - ❑ *Un **cookie ou témoin de connexion** est un fichier texte enregistrée sur le disque dur de l'internaute par le navigateur*
 - ❑ le cookie d'identification permet de naviguer entre les différentes pages d'un site en restant identifié ;
 - ❑ le cookie de préférence permet de mémoriser la langue dans lequel la page doit s'afficher ;
 - ❑ le cookie publicitaire permet de proposer des publicités ciblées.
-

