

4-Les sources virtuelles (bibliothèques virtuelles)

- Une bibliothèque numérique (ou en ligne ou électronique) est un ensemble d'écrits numérisés et accessibles à distance (en particulier via Internet¹).
- Les sources virtuelles quant à elles sont des écrits numérisés et accessibles à distance via l'Internet ou l'intranet.

4.1. Rappel sur l'internet et l'intranet

L'**Internet** permet également d'envoyer des courriers électroniques, de faire du partage de fichiers, etc.

Le **web** (World Wide Web) est un système hypertexte public fonctionnant sur **Internet**, dont il n'est qu'une application.

Les **moteurs de recherche** fournissent un grand nombre de résultats, qui ne sont pas toujours pertinents et n'indexent qu'une petite partie des documents disponibles sur le web.

Le **web invisible**, concerne les documents non indexés par les moteurs de recherche.

Les ressources du **web invisible** comprennent :

- les sites web construits autour d'une base de données (interrogeable 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 pages écrites dans des formats propriétaires, et
- les intranets.

Moteurs de recherches

Les moteurs de recherche sont les outils d'accès au web les plus utilisés. Le moteur de recherche fonctionne sur un système radicalement différent de celui de l'annuaire. Des robots logiciels (appelés **crawlers** ou **spiders** ou **bots**) parcourent le Web de lien en lien et indexent automatiquement les documents trouvés, permettant ainsi une recherche par mots-clés. Les index des moteurs de recherche mondiaux indexent plusieurs milliards de pages web.

Utilisation

Les recherches dans un moteur se font à l'aide de mots-clés. Ceux-ci peuvent être nombreux et combinés à l'aide des opérateurs booléens (voir section 4.3).

Les moteurs de recherche proposent la possibilité d'effectuer des recherches simples et avancées. Ces dernières offrent d'importantes possibilités pour paramétrer la recherche.

Le web a été inventé par Tim Berners-Lee en 1989 pour le CERN (organisation européenne pour la recherche nucléaire) et s'est démocratisé dans les années 1990.

Avantages des moteurs de recherches

- La possibilité de faire des recherches dans une grande masse d'informations.
- Le fait d'obtenir rapidement des informations précises sur des sujets divers et variés.
- Le fait de pouvoir trouver des documents spécifiques.
- Le fait de repérer des sites récents ou ayant été écartés des annuaires.
- Le fait de pouvoir réaliser des recherches complexes en utilisant la logique booléenne.

Inconvénients des moteurs de recherches

- Absence de contrôle des informations (présence d'URL périmées dans les résultats).
- Des interrogations qui semblent complexes car les interfaces changent d'un moteur de recherche à l'autre.
- Des résultats parfois surprenants car l'indexation est automatique.
- Des résultats qui sont classés selon un ordre qui donne la priorité à la popularité des Informations.

¹ Wikipédia

Les moteurs de recherche spécialisés

L'objectif des moteurs de recherche spécialisés dans un domaine disciplinaire est de pallier les inconvénients des outils généralistes.

- [Google Scholar](#), littérature de recherche universitaire
- [Scirus](#), information scientifique (Elsevier)
- [Wikipedia](#)
- [Wikiseek](#), recherche dans Wikipedia
- [Google Patents](#), brevets
- [In-extenso](#), sciences humaines et sociales
- [MedHunt](#), information médicale
- [PhiBot](#), information scientifique
- [Science Research](#), sciences
- [Scitopia](#), sciences et technologies
- [Web 2.0 Search Engine](#)

Les métamoteurs ou métaindex

Outils créés plus récemment que les précédents, les métamoteurs interrogent plusieurs outils de recherche simultanément. Les métamoteurs peuvent être consultés et utilisés directement sur le web ou installés en client sur votre poste. Les métamoteurs logiciels permettent d'envisager un travail de veille avec la gestion de profils et l'établissement de calendriers d'exploration.

Utilisation

Ces outils n'ont pas d'index en propre mais utilisent les index d'autres outils de recherche (annuaires, moteurs de recherche, métamoteurs, encyclopédies, bases de données). Ils éliminent les doublons et affichent les résultats selon des critères de pertinence.



4.2. Les bases de données

Les typologies de documents

Nature de l'information :

- **Documents textuels** : tous les documents où prédominent le texte ;
- **Documents non-textuels** : document où l'image et/ou le son prédominent.

Niveau d'information :

- **Les documents primaires ou originaux** : comportant l'information brute, c'est-à-dire lue ou vue par le lecteur dans le même état où l'auteur l'a écrite (ex: livre) ;
- **Les documents secondaires** : Comportent des données signalétiques ou analytiques sur des documents primaires (notice de description catalographique, indexation, résumé...) Visent à identifier, décrire et rechercher les documents primaires. Ex. : les bibliographies, index.

Reconnaître, et se repérer

- Savoir reconnaître les différents types de documents ;
- Savoir reconnaître les différents genres documentaires ;
- Savoir identifier un document particulier ;
- Savoir se repérer dans la structure propre à chaque type de document.

Type de publication :

- **Monographie ou Ouvrage** : document formant un tout, en un ou plusieurs volumes, soit qu'il paraisse en une seule fois, soit que sa publication s'étende sur une durée limitée selon un plan établi à l'avance. La monographie s'oppose aux publications en série et est identifiée par un numéro ISBN ;
- **Publication en série ou le périodique** : paraît en fascicules ou volumes successifs, s'enchaînant numériquement ou chronologiquement pendant une durée non limitée à l'avance. C'est le cas des revues, journaux, annuaires, séries de comptes rendus, rapports d'institutions et de congrès, et des collections de monographies. Elles sont identifiées par un numéro ISSN ;
- **Littérature "blanche"** : documentation accessible, commercialisée, largement diffusée (ex : les revues scientifiques, les livres, les banques de données scientifiques...) ;
- **Littérature "grise"** : documentation non commercialisée, diffusée de manière restreinte (ex : actes de congrès, thèses, rapports...).

Structure propre à chaque type de document

L'organisation d'un livre est différente de celle d'un périodique. Un site web n'est pas structuré comme une base de données ou un document linéaire. Une microfiche ne se "lit" pas comme une vidéocassette, etc.

Les principaux éléments à connaître pour se repérer dans un **livre** sont :

1. la couverture
2. la page de titre
3. Introduction, préface, avant-propos
4. La table des matières
5. Partie finale du livre (ou annexes)
6. L'achèvement d'imprimer

Les principaux éléments à connaître pour se repérer dans un **périodique** (une revue) sont :

1. La couverture
2. La page de titre (qui peut être très différente de la couverture)
3. L'ours (l'endroit de la revue où sont mentionnées toutes les informations éditoriales)
4. Le sommaire du numéro
5. Le dos (sur lequel figurent souvent la numérotation et la date de parution)
6. L'ISSN

Catalogue et localisation

Le catalogue référence tous les fonds d'une bibliothèque. Il en décrit en outre les enregistrements sous une forme physique (format, nombre de pages, etc.) et intellectuelle (mots sujets qui caractérisent la ressource). Ces données qui identifient le document et le décrivent physiquement et intellectuellement constituent **une notice**.

Le catalogue est une **base de données** un peu particulière dans la mesure où quand les bases de données au sens strict vous fait accéder le plus souvent à un article en ligne ou bien aux références de cet article, un catalogue vous permet **d'accéder à un document présent sur un rayon via une adresse codifiée** : la cote (ex. S 150.195 DRW = ouvrage conservé en salle S, rayon Evolution, au milieu des autres ouvrages de Darwin).

Bases de données

Une base de données au sens où nous l'entendrons dans ce guide est un ensemble d'informations stockées dans un réservoir structuré permettant de mettre ces informations en relation les unes avec les autres.

L'université, à travers ses laboratoires et son service commun de documentation produit un certain nombre de bases de données ; un certain nombre d'autres bases de données se trouvent en accès libre sur le web.

Avec l'extension du web marchand, les bases de données accessibles par abonnement deviennent de plus en plus nombreuses et incontournables, alors même que le nombre de leurs éditeurs, au gré des fusions et des acquisitions devient plus restreint.

Les positions dominantes de ces éditeurs sur le marché rendent les abonnements à ces bases de données de plus en plus coûteux pour les universités.

Malheureusement, le marché francophone ne s'étend pas encore à la même vitesse que le marché anglophone, c'est pourquoi le nombre de BDD en français n'atteint pas encore celui des BDD en anglais ou dans d'autres langues que nous pouvons vous proposer.

4.3. Démarches d'exploitation des bases de données

Comment Trouver une source ?

Avant de faire une recherche, que ce soit dans un catalogue de bibliothèque ou dans Internet, il convient de bien cerner son sujet et de choisir des mots-clés significatifs qui représentent les concepts sur lesquels on veut se documenter. Pour faire une recherche efficace, c'est aussi une bonne idée de tirer parti des fonctions avancées des outils de recherche.

Ainsi, dans une recherche en bibliothèque ou sur le Web, il est très utile de recourir aux **opérateurs logiques (ET, OU, SAUF, etc.)** pour combiner les mots clés, ainsi que de limiter la recherche par langue, par date de publication ou par type de document, etc. Dans un catalogue de bibliothèque, on peut aussi utiliser des fonctions comme la recherche par mot-clé, par sujet ou par vedette-matière, ou encore la recherche par titre ou par auteur, etc.

Les index (ou lexique) et les thésaurus

L'indexation est la phase de l'analyse documentaire qui permet d'affecter au document un ou plusieurs termes significatifs. Il existe plusieurs outils documentaires pour choisir les termes significatifs :

- l'index ou le lexique de descripteurs (on parle également de mots-clefs) : ce sont des listes alphabétiques de termes contrôlés. Le documentaliste-indexeur tirera de cette liste les descripteurs qui reflètent le mieux les concepts de l'article.
- si les descripteurs sont organisés de manière hiérarchique (terme générique décliné en plusieurs termes spécifiques) avec des renvois d'équivalence -synonyme de- et d'association -voir aussi -, on parle de thésaurus. Ex : le Mesh® utilisé dans Medline®.

Exemple d'Index ou Lexique (MeSH en français, INSERM) :

ABCES ABDOMINAL
ABCES CERVEAU
ABCES MUSCLE PSOAS
ABCES PARODONTAL

Exemple de Thésaurus :

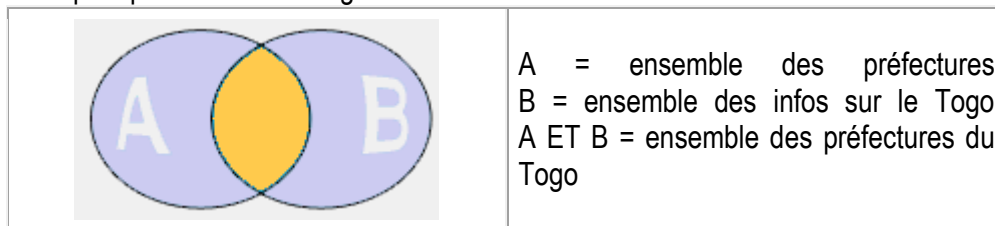
Terme générique	ABCES
Termes spécifiques	...ABCES ABDOMINAL
	ABCES FOIE
	ABCES AMBIEN FOIE
	...ABCES CERVEAU
	...ABCES MUSCLE PSOAS
	...ABCES PARODONTAL

Les opérateurs logiques booléens

- Il existe des outils documentaires qui permettent d'affiner, d'orienter une recherche afin d'éviter le "bruit" (l'obtention de données hors sujet lors d'une interrogation) ou, au contraire le "silence" (l'obtention d'un nombre insuffisant de notices par rapport au nombre pertinent de réponses qu'on aurait dû obtenir).
- L'usage des **opérateurs booléens ou logiques** (ET, OU, SAUF) est indispensable pour lier des termes à rechercher et écrire des équations de recherche selon une logique mathématique (algèbre de Boole).

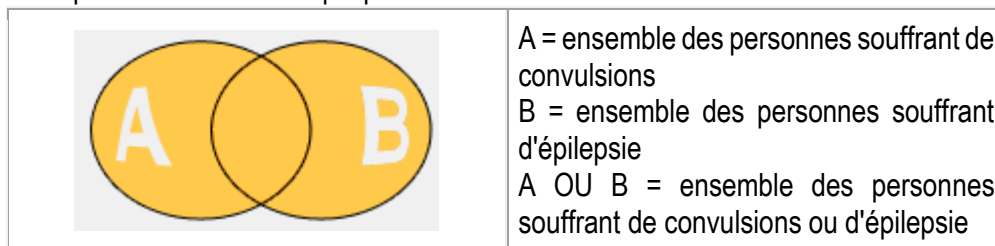
- **ET/AND (intersection)** : pour obtenir les notices comportant la présence simultanée des mots-clefs recherchés.

Exemple : préfectures ET Togo



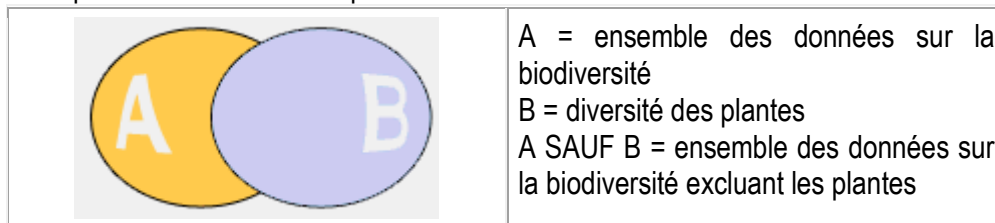
- **OU/OR (union)** : pour obtenir les notices comportant au moins l'un des descripteurs recherchés.

Exemple : convulsion OU épilepsie



- **SAUF/NOT (exclusion)** : sert à éliminer les notices incluant le mot-clef rejeté.

Exemple : Biodiversité SAUF plantes



Les opérateurs de proximité

Les **opérateurs de proximité** (ADJ, NEAR, NEARx, WITH, FOLLOWED BY, SAME selon les bases de données) sont des opérateurs qui permettent de retrouver deux termes dans un même champ ou une même phrase. Certains opérateurs sont assez précis pour que l'on puisse choisir le nombre de mots entre les deux termes recherchés et l'ordre de ces mots.

- les opérateurs de proximité sont utilisés essentiellement pour la recherche en texte libre dans les champs Titre et Résumé.

*Ex : la requête "**FDS/UL** NEAR **BPA** in AB" retrouve les notices qui contiennent **FDS/UL** et **BPA** dans le champ Résumé sans notion d'ordre.*

- ils sont conseillés pour trouver les expressions.

Ex : la requête "Micro ADJ Ordinateur" retrouve les notices qui contiennent le terme Micro suivi immédiatement d'Ordinateur (très utile pour les expressions).

*Ex : la requête "**Guide** NEAR3 **Biologie**" retrouve les notices qui contiennent les deux termes dans une limite de trois mots consécutifs dans une même phrase.*

- les opérateurs de proximité peuvent être aussi employés pour les recherches sur les noms d'auteur qui sont souvent entrés avec des variantes.

Ex : la requête " Smith NEAR3 J" retrouve John Smith, John Scott Smith, John S Smith etc.*

- Attention : le nom et le mode de fonctionnement exacts de ces opérateurs varient d'une base à l'autre. Il est conseillé de consulter le mode d'emploi de la base de données interrogée.

Les opérateurs de fréquence

- Certaines interfaces d'interrogation proposent un opérateur de fréquence (FREQn) pour spécifier le nombre d'occurrence d'un terme dans un champ.

C'est très utile pour la recherche en mots libres dans le champ Résumé : *pression sanguine.ab/freq=5* pour rechercher les notices dont le résumé (ab) comporte au moins 5 fois le terme "pression sanguine".

Les priorités entre opérateurs

- Si vous utilisez plus d'un opérateur booléen ou de proximité dans une requête, sachez qu'ils opèrent dans un ordre pré-établi qui varie selon les bases.

- de manière générale, l'ordre des opérateurs booléen est le suivant :

OR

AND

NOT

Et ils opèrent de la gauche vers la droite

- les opérateurs de proximité peuvent s'intercaler entre les opérateurs booléens.

Ex : *pain NEAR20 morphine AND ganglia OR tumor lesion*

1. Le système commence par chercher les articles contenant "ganglia", "tumor" ou "lesion" (OR a priorité)

2. Puis, il cherche les articles contenant le terme "pain" à au maximum 20 mots de "morphine" (NEAR a priorité devant AND)

3. Ensuite, seuls les articles contenant à la fois les deux étapes précédentes sont sélectionnées (AND opère en dernier)

4. Dans la plupart des bases de données, il est remplacé par l'espace. *Lomé AND Togo* = je veux tous les résultats qui contiennent le mot-clé Lomé et à la fois le mot-clé Togo.

- pour changer les priorités entre opérateurs, vous pouvez utiliser des parenthèses. Les opérateurs entre parenthèses opèrent avant ceux situés en dehors des parenthèses.

Ex : *(genes OR chromosomes NEAR5 splicing) OR cloning*

1. Le système recherche tout d'abord les articles contenant "genes" ou "chromosomes" quand ils sont situés à 5 mots de splicing.

2. Ensuite, il cherche les articles "cloning"

3. **Togo OR Ghana** ramènera des pages sur les deux pays

Les troncatures

- L'usage des troncatures permet en remplaçant un ou plusieurs caractères de trouver les mots au singulier ou au pluriel, d'autres mots de la même famille. (ex : *robot** retrouve robot, robots, robotique, robotisé).
- Les symboles varient selon les logiciels : *, ?, +, \$.
- Certains logiciels font la différence entre une troncature illimitée, qui remplace aussi bien un seul caractère qu'une chaîne de caractères et une troncature limitée à un seul caractère (wildcard, joker). (ex. de troncature illimitée : *cat** retrouve catatonique, cats... ; et ex. de troncature limitée : *M?cdonald* retrouve à la fois Macdonald et Mcdonald).

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.

4.4. Évaluer une source

Évaluer la pertinence et la fiabilité de l'information

Évaluer la pertinence d'une source implique de déterminer si les informations qu'elle contient correspondent au sujet de recherche et répondent à l'objectif poursuivi.

En effet, les documents imprimés comme **les encyclopédies, les dictionnaires, les livres et les revues scientifiques** sont publiés à la suite d'un long processus de validation auquel participent divers spécialistes (auteurs, réviseurs de contenu, réviseurs linguistiques, graphistes, etc.), s'assurant ainsi de la fiabilité de l'information présentée.

Cependant

Les documents trouvés dans Internet, quant à eux, peuvent être classés dans deux grandes catégories :

- ceux qui sont publiés par des sources fiables (**gouvernements, organismes publics ou largement reconnus**, etc.) qui ont aussi recours à un long processus de validation;
- ceux qui sont publiés par des sources dont la crédibilité reste à établir (**sites Web personnels, blogs, réseaux sociaux, etc.**).

Dès lors, il revient à la personne qui consulte ces sources d'information d'en établir la fiabilité.

Évaluation de la pertinence en consultant plusieurs sources

Pour évaluer la pertinence et la fiabilité d'une source d'information et de chaque information qui s'y trouve, il convient :

- d'évaluer si l'information trouvée correspond bien au sujet de recherche;
- de corroborer l'information trouvée en consultant plus d'une source;
- d'examiner soigneusement l'ensemble de la source;
- de poser un regard critique sur l'information.

Évaluation de la fiabilité en consultant plusieurs sources

Il faut absolument éviter de ne consulter qu'une seule source d'information (à remarquer que c'est aussi vrai pour les sources plus « officielles ») et, au contraire, recommander aux étudiants de vérifier dans au moins trois sources si l'information qu'ils ont trouvée est corroborée dans celles-ci.