

université
PARIS-SACLAY

UE 16 PIX

2e A – 2023-2024

Volet Recherche
d'information

Université Paris-Saclay : Les Bibliothèques
2023-2024

Mis à jour le : 02/11/2023

Objectifs de la séance :

- Pouvoir trouver la bonne information sur Internet et les ressources académiques
- Se constituer sa propre veille d'information

UE 16 : Initiation aux pratiques professionnelles

Initiation à la recherche documentaire et à l'informatique dans le cadre du PIX

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

La séance d'aujourd'hui fait partie de l'UE 16 coordonnée par M. Thomas CANDELA qui comprend une initiation à la recherche documentaire et à l'informatique.

Cette séance est destinée à vous accompagner dans le volet recherche documentaire du PIX (1).

(1) Toute question relative au PIX doit être posée à votre responsable de groupe.

Plan de la formation

1. Rechercher de l'information
2. Évaluer les résultats d'une recherche
3. Référencer une ressource
4. Veiller l'information

université
PARIS-SACLAY

I – Rechercher de l'information



Université Paris-Saclay : Les Bibliothèques
2023-2024

Le Lumen : plus qu'une simple bibliothèque



Source : <http://www.beaudouin-architectes.fr>

5 Formation Information Scientifique et Technique - Lumen
2023-2024

université
PARIS-SACLAY

Le Lumen est plus qu'une simple bibliothèque.

Il réunit des fonds issus

- de Centrale Supélec,
- de l'École normale supérieure Paris-Saclay
- et de l'Université Paris-Saclay (notamment les fonds de la bibliothèque de pharmacie, et une partie des collections d'Orsay),

Il propose des millions de livres, revues et articles, en papier ou en ligne, couvrant tous les domaines d'activité du plateau du Moulon : ingénierie, biologie, pharmacie, chimie, physique, sciences humaines et sociale.

Du wifi, des copieurs et imprimantes, des casiers pour recharger son matériel informatique

+ accompagnement de projets innovants ; expositions, conférences, ateliers et autres événements

Des espaces de travail et de loisirs adaptés aux nouveaux usages : travail individuel, grands espaces de travail collectif, lieux de rencontre, etc.

-> Une sélection d'ouvrages en libre accès vous attend à l'accueil de la bibliothèque de Centrale Supélec, en attendant l'ouverture définitive de nos locaux.

-> Les ressources 24h./24h. sont, quant à elles, continuellement à votre disposition,

quelque soit le lieu ou l'heure de votre connexion : <https://www.bibliotheques.universite-paris-saclay.fr/explorer-les-ressources/ressources-24/24>

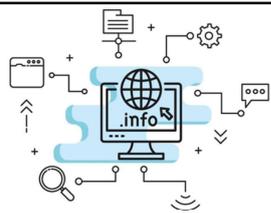
-> Ces ressources sont aussi accessibles via votre catalogue Focus :

https://upsaclay.focus.universite-paris-saclay.fr/primos-explore/search?vid=33UDPS_VU1&lang=fr_FR

-> N'oubliez pas la solution Lean Library pour accéder directement aux ressources électroniques de Paris-Saclay via PubMed, Google Scholar ou les sites des éditeurs !

<https://www.bibliotheques.universite-paris-saclay.fr/explorer-les-ressources/library-access>

Nous suivre



- Réseaux sociaux

@lumen.upsaclay

<https://www.universite-paris-saclay.fr/lumen>

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

Avant de venir à la bibliothèque, n’hésitez pas à consulter les réseaux sociaux (Instagram) pour vérifier l’ouverture de la bibliothèque ainsi qu’**Affluences** pour les horaires (accès aussi à **Affluences** sur la page d’accueil des BU <https://www.bibliotheques.universite-paris-saclay.fr/> et **l’appli téléchargeable pour smartphones et tablettes**).

Consulter les réseaux sociaux vous évitera un déplacement inutile en cas de fermeture indépendante de notre volonté.

Prenez le réflexe de consulter aussi votre mail Paris-Saclay : vous serez ainsi informés de toute l’actualité du campus.

L’application **Affluences** vous permet également de réserver une salle de travail en groupe.

Un réseau de bibliothèques



Lumen learning



Kremlin-Bicêtre



BU Sciences et STAPS à Orsay.



BU droit éco gestion à Sceaux.

8 documents pour 3
semaines sur tous les
sites
+
La navette

7

Formation Information Scientifique et Technique -
Lumen 2023-2024

université
PARIS-SACLAY

Prêt sur tous les sites : 8 doc pour 21 jours pour le niveau L2 avec une possibilité de prolongation de 15 jours.

Votre statut vous donne accès à toutes les BU de Paris-Saclay, y compris jusqu'à Evry et Versailles-Saint-Quentin.

Un service de navette vous est proposé. Vous pouvez vous faire parvenir un document provenant de l'une des bibliothèques présentes sur la diapo, en choisissant votre bibliothèque de retrait (le service de navette ne fonctionne pas avec les BU d'Evry ou de Versailles-Saint-Quentin).

Pour ce faire, vous pouvez effectuer une demande à l'intérieur de la fiche du document en question dans notre catalogue Focus (bouton « demande ») ; vous pouvez également effectuer cette demande en direct à l'un.e des bibliothécaires posté.e.s à l'accueil de la bibliothèque de votre choix.



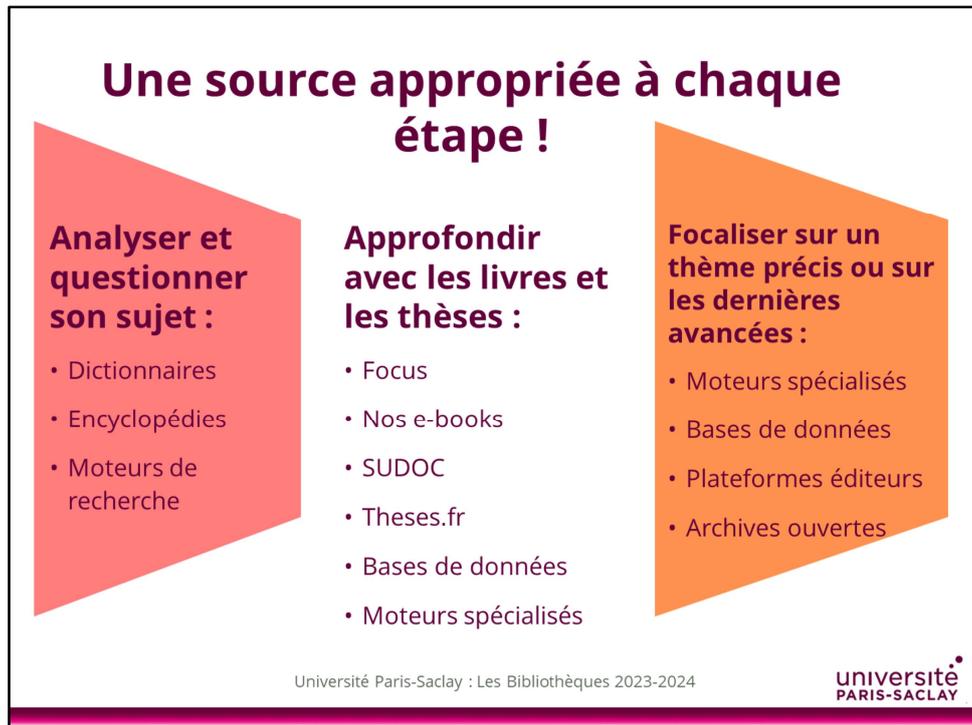
Une petite idée de ce qu'il y a derrière ces logos ?

- Lilo est ce qu'on appelle un « métamoteur » : il explore à votre place d'autres moteurs de recherche ; sa particularité est qu'il vous permet de choisir des projets sociaux et environnementaux à soutenir : chacune de vos recherche dans Lilo contribue au financement d'une ONG ; Lilo ne collecte pas les données personnelles des utilisateurs.
-  DuckDuckgo : autre métamoteur. Il ne collecte pas vos données personnelles.
- Google : le leader des moteurs de recherche, mais beaucoup de publicité et de résultats commerciaux parfois peu pertinents.
-  Qwant : Concurrent français de Google il entend comme DuckDuckgo être respectueux de vos données personnelles.
-  Mozilla Firefox, lui, est un navigateur libre et gratuit.

Au fait, savez-vous bien faire la différence entre navigateur et moteur de recherche?

Sinon, voir : <http://www.e-dilik.fr/web/difference-entre-un-navigateur-et-un-moteur-de-recherche/> (billet de blog ancien mais clair)

- EM-Premium est un éditeur en ligne auquel vous avez accès par la BU.
- Pubmed est la base de données bibliographique de la NLM (Bibliothèque nationale de médecine des Etats-Unis) : elle signale essentiellement des articles scientifiques dans le domaine médical.
- Le Sudoc est un catalogue national français en ligne qui référence les ressources de toutes les Bibliothèques universitaires françaises.
- Wikipedia est une encyclopédie en ligne gratuite et collaborative.
- NetVibes est un outil de veille en ligne par flux RSS (cf. partie sur la veille)



La recherche documentaire est une démarche progressive qui se réalise étape par étape.

A chacune des étapes correspondent des démarches et des outils adaptés.

Il convient d'abord d'analyser et de questionner son sujet.

On définit les termes et on cerne son périmètre. Pour cela, on a notamment recours à des ressources généralistes qui sont :

Dictionnaires

Encyclopédies

Moteurs de recherche (à utiliser avec discernement)

C'est une étape à ne pas négliger, afin de **poser de solides fondations** à tout le travail qui va suivre.

Il sera bon de s'y reporter constamment. **La démarche documentaire est itérative.**

Conseil précieux : **Tenir un journal de bord** qui va vous suivre à chaque pas.

Rechercher de l'information est une démarche beaucoup plus complexe. La recherche documentaire est une démarche progressive : elle s'effectue par étapes successives. Il ne faut pas foncer « tête baissée » mais commencer par réfléchir aux objectifs que doit atteindre cette recherche.

Pour avoir une vue d'ensemble sur un sujet ou des connaissances approfondies :

un livre,

une thèse

En revanche, pour connaître les dernières avancées sur un sujet de recherche, il va falloir notamment identifier des articles scientifiques.

Comment chercher ?

Définir précisément ses mots-clés

Exemple de sujet:

Effets secondaires du vaccin contre le papillomavirus

- Termes **du sujet** : mots clés directement pris dans l'intitulé du sujet. Ex. **Effet secondaire**
- Termes **plus larges** : mots clés plus larges
Ex. **Complication**
- Termes **plus précis** : mots clés plus précis
Ex. **Effet indésirable à long terme**
- **Synonymes** : mots de même sens. Ex. **Effet indésirable**
- **Traductions** en anglais. Ex. **Side effect**

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

L'importance du choix des mots-clés

1) Il n'est pas sûr que nous-mêmes ayons une appréhension globale et exhaustive du sujet, surtout en début de travail. Ce n'est **jamais** le cas.

Multiplier les mots clés, explorer à fond chaque concept et chaque définition permet d'y voir clair mais aussi de poser des limites.

-> **Poser des limites est essentiel** :

a) Pour le travail de recherche documentaire lui-même

b) Et mine de rien aussi pour votre travail final : toute réflexion sur un sujet commence par sa définition et sa circonscription par des limites.

2) **Il n'est pas sûr que les auteurs aient employé les mêmes termes** ou aient exactement traité des mêmes concepts que vous.

Pour compliquer la chose, les moteurs de recherche, les bases de données, les catalogues, les sites d'éditeurs (ou autres) ajoutent eux-mêmes manuellement ou automatiquement des mots-clés qui ne seront pas forcément les vôtres.

Autant dire qu'en face, vous devez être convenablement armés.

3) Et si une liste de mots-clés pertinente est un arsenal efficace, il convient de se souvenir que la langue de la recherche scientifique est l'Anglais.

- Termes importants du sujet (parfois appelés « termes significatifs »)

ATTENTION : retirez systématiquement les mots « vides » tels que déterminants (le, la, les, d', des, etc.) ou prépositions (dans...).

Pensez aussi aux synonymes !

- Termes plus larges (appelés aussi « génériques ») : à utiliser quand vous n'avez pas assez de résultats.
- Termes plus précis (appelés aussi « spécifiques »): à utiliser quand vous en avez trop.

Pensez aux traductions des termes en anglais (=langue de la science et de beaucoup de bases de données scientifiques et médicales comme PubMed).

Tableau des mots-clés

Sujet : Effets secondaires du vaccin contre le papillomavirus

Termes significatifs	Termes plus larges	Termes plus précis	Synonymes
Effet* secondaire	Pharmacovigilance Risque Effet* Complication* Contre-indication*	Effets secondaires métaboliques des médicaments et des substances Effets indésirables à long terme	Effet* indésirable* Toxicité* Nocivité*
Vaccin*	Prévention Immunothérapie active Vaccinologie	Couverture vaccinale Vaccination de masse Gardasil Cervarix	Vaccination
Papillomavirus	MST Maladie sexuellement transmissible Virus		Papillomaviridae VPH Virus du papillome humain HPV Human Papillomavirus

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

Nous vous proposons de mettre au clair votre liste de mots-clés sous la forme d'un **tableau** (n'oubliez pas les mots en anglais, le cas échéant!).

Combinez les cases entre elles comme vous l'entendrez pour enrichir votre recherche.

Ce tableau est évolutif : vous trouverez d'autres mots-clés au fil de votre recherche et peut-être même que vous en éliminerez qui ne sont pas pertinents !

Pour vous aider à trouver les bons mots-clés, vous pouvez par exemple vous servir de Loterre (<https://skosmos.loterre.fr/fr/>), de « Vitrine linguistique » (<https://vitrinelinguistique.oqlf.gouv.qc.ca/>), de HeTOP (<https://www.hetop.eu/hetop/fr/?q=&home>) ou d'AcadPharm (cf. diapo 13) : les trois vous seront utiles aussi pour les traductions de termes en anglais.

Pour traduire un terme ou un texte entier, DeepL sera aussi utile : <https://www.deepl.com/fr/translator>

Notez que quoique performant, DeepL n'est pas d'une rigueur extrême pour la traduction des termes scientifiques pointus. **Dans ce cas, recourez plutôt à la version anglaise d'une page Wikipedia (à partir de la page française sur votre sujet) ou à la bibliographie d'un article de Sciences de l'Ingénieur.** Exemple, la notion de « doublet non liant » en chimie : DeepL donne de mauvaises traductions tandis que la version anglaise de la page Wikipedia « Doublet non liant » donne la bonne.

Petite parenthèse écoresponsable : Lorsque vous tapez une requête dans un moteur de recherche Internet, cela fait travailler de nombreux serveurs distants situés partout dans le monde et souvent alimentés par une électricité provenant des énergies fossiles. L'impact du numérique et en particulier

d'Internet sur l'environnement est donc non négligeable.

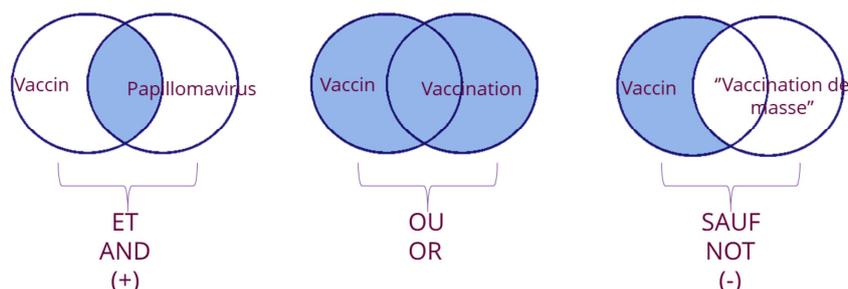
Les bons réflexes :

- Ne pas utiliser systématiquement les moteurs de recherche pour trouver un site. Enregistrer les ressources que l'on visite souvent dans ses favoris ou marque-pages. Ecrire ou coller l'adresse (URL) des sites voulus dans la barre d'adresse.
- **Avoir une sélection judicieuse de mots-clés !** (cela limite le travail des serveurs distants et polluants!)

Voir : <https://www.apc-paris.com/actualite/20-eco-gestes-a-adopter-pour-numerique-responsable>

Comment chercher ?

Utiliser les opérateurs booléens



Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

Combiner les mots clés (1)

Il est également utile de combiner les mots clés à l'aide des opérateurs logiques ET / OU / SAUF pour obtenir des équations de recherche pertinentes :

ET (« AND » en anglais ; « + » devant un terme dans Google) : pour obtenir des documents comportant l'un et l'autre mot (intersection). L'opérateur ET est souvent l'opérateur par défaut.

OU (« OR ») : pour obtenir des documents comportant l'un **ou** l'autre des mots (addition). Le OU est très utile pour combiner notamment les termes synonymes ou de sens voisin. Attention : risque d'avoir trop de résultats, à utiliser donc avec précaution.

SAUF (« NOT » ; signe « - » devant un terme dans Google) : pour obtenir des documents excluant un mot (exclusion)

Utiliser les troncatures

La troncature est un signe, souvent placé à la fin d'un mot, qui remplace un ou plusieurs caractères. Elle permet de lancer une recherche sur toutes les variantes d'un mot ayant une racine commune et d'élargir ainsi votre recherche.

Exemple : soign* = soignant, soignants, soigné, soignés, soigner, etc.

Vous pouvez utiliser des troncatures dans les plupart des outils de recherche. La troncature est souvent symbolisée par * (voir diapo 12) mais parfois aussi par : !, %, ?

Attention : certains moteurs de recherche ne comprennent pas la troncature.

UTILE: Mettre une expression (même un seul terme) entre guillemets force le moteur de recherche à trouver la formulation exacte ; exemple : **"vaccination de masse"**

(1) Vous pouvez consulter avec profit la page suivante : https://methodoc.univ-littoral.fr/wp-content/uploads/opale/V2.%20Les%20astuces%20de%20recherche/co/Les_operateurs_booleens.html de l'Université Littoral Côte d'Opale (Ulco).

Dictionnaire et Encyclopédie

- Dictionnaire de l'Académie Nationale de Pharmacie (Acadpharm)



- Wikipédia



Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

Pour avoir une définition précise ou des informations synthétiques sur un sujet, pour débiter une recherche et la délimiter, **dictionnaires et encyclopédies** sont un bon point de départ.

Le dictionnaire de l'Académie Nationale de Pharmacie : Acadpharm

<https://dictionnaire.acadpharm.org/w/Acadpharm:Accueil>

Il rassemble les définitions sur les médicaments et les produits de santé, les sciences physico-chimiques, la biologie, la santé publique, l'environnement et la santé, la diététique, la nutrition, la cosmétologie. Sont également décrits les symptômes majeurs des principales pathologies.

Les définitions sont rédigées par des spécialistes des différents domaines, dont font partie certains chercheurs de l'université Paris-Saclay.

Wikipédia

<https://fr.wikipedia.org/>

Créée en 2001, c'est une encyclopédie gratuite, collaborative, établie sur Internet, universelle, multilingue. Chaque lecteur peut aussi modifier une page. Elle peut aussi être copiée, utilisée et vendue librement, tant que la source et les auteurs sont mentionnés.

Les informations sont soumises à une surveillance constante mais sans système de validation par des experts. **L'exactitude du contenu n'est donc pas garantie.**

Wikipedia

- Encyclopédie gratuite et collaborative.
- Pas de validation par des experts.
- Un outil à maîtriser.



Université Paris-Saclay : Les Bibliothèques 2023-2024 

POINTS PIX:

- Chaque article de Wikipédia a une page principale, par exemple celle de l'article « Paracétamol »: <https://fr.wikipedia.org/wiki/Parac%C3%A9tamol>
- Mais comme Wikipédia est une encyclopédie collaborative/participative, les pages évoluent dans le temps ; on peut donc retrouver l'historique de ces évolutions dans l'onglet « Voir l'historique »
- On peut naviguer ainsi par date dans les différentes versions (privilégier classement par 500 versions les plus récentes ou anciennes et faire un Ctrl+F pour trouver la date désirée, par exemple : **Ctrl+F** et rechercher **6 mai 2020**)
- Vous pouvez remarquer que chaque auteur de modification est identifié par un pseudonyme (utilisateur inscrit avec un compte Wikipédia) ou par son adresse IP (1) (utilisateur anonyme)
- Il est toujours important de connaître l'auteur d'une information, **sur Wikipédia ou ailleurs**. Cependant, il est devenu difficile de savoir qui se cache derrière une adresse IP (2)
- Les modifications de Wikipédia sont contrôlées par des utilisateurs que l'on appelle « patrouilleurs » ; certains sont des robots virtuels : leur pseudonyme se finit par « -bot »
- Les modifications mineures sont marquées d'un « **m** » en gras.

-> Nous attirons votre attention sur les articles « de qualité » signalés par une étoile de couleur or et les « bons articles » (étoile couleur argent)

-> Voir aussi les portails thématiques (<https://fr.wikipedia.org/wiki/Portail:Accueil>) ; les plus intéressants pour vous : <https://fr.wikipedia.org/wiki/Portail:Médecine> et <https://fr.wikipedia.org/wiki/Portail:Pharmacie>).

Dans le portail « Médecine », remarquez l'étoile d'or de la version anglaise du portail, tandis que la française a une étoile d'argent : eh oui, l'anglais est la langue de la science.

(1) Une adresse IP est un identifiant numérique composé de nombres séparés par des points. Chaque dispositif informatique ayant accès à Internet possède une IP : votre box, votre ordinateur, votre tablette, votre téléphone portable ou même votre console de jeux. Ainsi, tout utilisateur est identifiable par son IP sauf s'il passe par un proxy ou un VPN. Les fonctions de navigation anonyme comme dans Firefox ou Chrome ne vous empêchent pas d'être identifiés. Elles créent juste un cache temporaire et vierge afin d'effacer vos informations personnelles une fois le navigateur fermé.

Voir : « Nouvelles méthodes de traçage en ligne : quelles solutions pour se protéger ? » <https://www.cnil.fr/fr/nouvelles-methodes-de-tracage-en-ligne-queelles-solutions-pour-se-proteger>

(2)« L'adresse IP est une donnée personnelle, encadrée par la CNIL » <https://www.silicon.fr/adresse-ip-est-une-donnee-personnelle-encadree-par-la-cnil-161972.html>

Wikipedia : les cadenas

 : semi-protection de la page
 : semi-protection longue
 : semi-protection étendue
 : protection


 WIKIPEDIA
The Free Encyclopedia

Tout le détail sur les différents niveaux de protection ici :
https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Niveaux_de_protection

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

Parfois, une page Wikipédia est l'objet d'une « guerre d'édition » ou bien de vandalisme. Les solutions sont le verrouillage total ou partiel des dites pages pour un temps plus ou moins long.

- semi-protection : mesure temporaire face au vandalisme et au « spam » des pages (pas de contributeurs anonymes avec adresse IP seule comme identifiant, seulement les utilisateurs « autoconfirmed » 4 jours après inscription : https://fr.wikipedia.org/wiki/Aide:Statuts_des_utilisateurs)
- semi-protection longue : sur les sujets sensibles (modification seulement par les « autoconfirmed »)
- semi-protection étendue : pour les sujets sensibles sous les feux de l'actualité en général (seulement les contributeurs « autopatrolled » : inscrits depuis + de 90 jours et avec 500 modifications à leur actif)
- protection : sur décision des administrateurs de Wikipédia (modification seulement par les administrateurs)

Les catalogues de bibliothèque

➤ Localiser et répertorier des documents

➤ Catalogue local / Catalogue national



Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

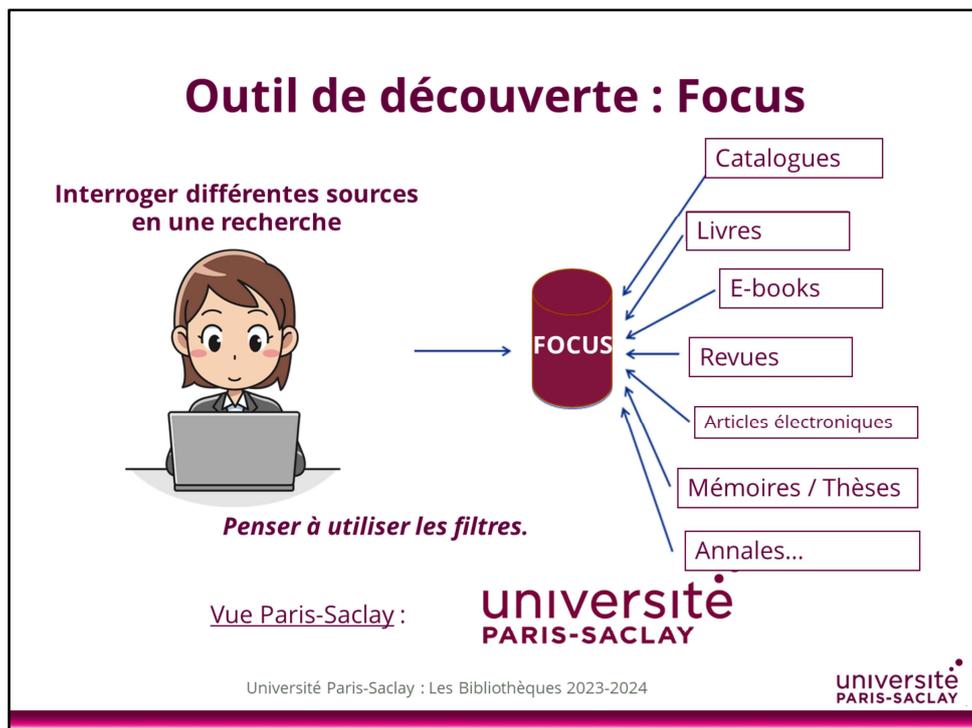
Une bibliothèque conserve des documents (livres, journaux, thèses ...).

Le catalogue permet de répertorier et de trouver ces documents par une recherche en ligne.

Il donne accès aux notices bibliographiques, c'est-à-dire à la description des documents : titre, auteur, éditeur, année de publication, mots-clés, résumé, etc.

Plus important, il indique si le ou les exemplaire(s) possédé(s) sont disponibles et à quel endroit (= notice d'exemplaire).

Il existe autant de catalogues que de bibliothèques mais il n'existe qu'un catalogue national pour toutes les BU : le Sudoc.



Focus est plus qu'un simple catalogue, c'est un « outil de découverte » qui permet de trouver via une seule et même interface de recherche toute la documentation proposée. L'outil est composé de plusieurs vues :

- Une vue consortiale contenant toute l'offre documentaire de l'ensemble des bibliothèques et centre de documentation de l'Université Paris-Saclay
- Une vue pour chaque institution composant de l'université Paris Saclay

On y trouve :

Les notices des livres et revues papiers
 Les notices des e-books et leurs accès
 Les notices des articles, revues, thèses en ligne avec leurs accès

Pour plus d'informations :

<https://www.universite-paris-saclay.fr/vie-de-campus/bibliotheques/guide-dutilisation-de-focus>

Accès à Focus :

La meilleure façon d'accéder à Focus est de s'y connecter par le bouton « Compte lecteur » sur [la page d'accueil des BU Paris-Saclay](#) (en haut à droite).

- Vue « Paris-Sud » : <https://universiteparissud.focus.universite-paris-saclay.fr/>
- Vue « consortiale » (tout Paris-Saclay) : <https://focus.universite-paris-saclay.fr/>

En cas de difficulté d'accès à nos bibliothèques physiques, le recours aux e-books peut être précieux.

Trouvez-les et lisez-les en y accédant par Focus.

Pour cela, utiliser les filtres suivants :

Disponibilité: accès en ligne

Type de ressource : Livres

Le SUDOC

Un catalogue commun aux :

- Bibliothèques universitaires françaises.
- Bibliothèques de certains établissements.



Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

« Le Système universitaire de documentation, en abrégé Sudoc, est un catalogue collectif alimenté par l'ensemble des bibliothèques universitaires françaises et de nombreux établissements documentaires de recherche. Il recense les documents en leur possession afin de permettre aux usagers d'effectuer des bibliographies, de localiser des ouvrages pour le cas échéant, effectuer des demandes de prêt entre bibliothèques (PEB).

« C'est l'Abes - Agence bibliographique de l'enseignement supérieur qui est chargée du maintien et du développement du Sudoc ainsi que de l'animation des réseaux de production de données bibliographiques. » (Source : Wikipédia)

Il est accessible à cette adresse : <https://www.sudoc.abes.fr/>

-> Si un document n'est pas présent dans les BU de Paris-Saclay, tout n'est donc pas perdu !

Une fois que vous avez trouvé ce que vous cherchez dans le Sudoc, cliquez sur « Où trouver ce document ? », ceci vous permettra non seulement de savoir où le trouver mais aussi s'il est disponible pour le PEB (ne contactez pas vous-même la BU concernée : ce sont vos bibliothécaires qui commanderont le document pour vous).

Vous trouverez tout sur le PEB ici : <https://www.bibliotheques.universite-paris-saclay.fr/le-pret-entre-bibliotheques-peb>



Encore quelques logos !

- **Yahoo!** est un moteur de recherche . « Yahoo! est une société américaine de service web appartenant depuis 2017 à Verizon Media » (Source : Wikipedia)
- **Bing** est le moteur de recherche développé par Microsoft.

-> Comme Google, ces deux moteurs de recherche vivent en grande partie de la publicité et privilégient souvent dans le classement des résultats de sources commerciales ou publicitaires.

Comme Google, ils recueillent vos informations personnelles et laissent des traces de votre navigation sur votre ordinateur, notamment avec les **cookies** (1). Officiellement, il s'agit de personnaliser les résultats ; en fait il s'agit plus souvent de vous faire parvenir de la publicité ciblée. Le règlement général sur la protection des données (RGPD) entré en vigueur le 25 mai 2018 a été mis en place pour limiter la politique intrusive des différents acteurs d'Internet à ce propos (voir : <https://www.vie-publique.fr/eclairage/19588-rgpd-reglement-general-sur-la-protection-des-donnees-de-quoi-sagit-il>)

- **Ecosia** : « Ecosia est un moteur de recherche solidaire allemand. Il reverse 80 % de ses bénéfices selon un programme de reforestation présent partout dans le monde » (source : Wikipédia). C'est un métamoteur qui utilise Bing, mais qui retraite ses résultats. La publicité lui sert à financer sa politique de reforestation. Voir : https://www.lemonde.fr/pixels/article/2020/01/19/ecosia-peut-on-vraiment-planter-des-arbres-en-utilisant-un-moteur-de-recherche_6026466_4408996.html
- Vous pouvez enrichir la sélection de métamoteurs de recherche présentée ici avec Startpage qui se dit aussi soucieux des données personnelles (bien qu'il propose de la publicité ou « annonces ») : <https://www.startpage.com> ; mais « en octobre 2019, Startpage BV est acquis par Privacy One Group, propriété de l'entreprise publicitaire System1. Cette acquisition questionne sur le respect de la vie privée. » (source : Wikipédia) Sur Internet rien n'est jamais tout blanc ou tout noir !
- Le métamoteur [Searx](https://www.searx.org/) offre des ressources alternatives mais il faut le paramétrer en mode un peu expert ([Préférences](#)). La sélection de l'onglet « Science » dans la recherche permet de trouver des articles scientifiques.

-> En fait, chaque moteur a ses atouts. Il convient de ne pas se contenter d'un seul et de les combiner en fonction de vos besoins.

(1) Un cookie est un fichier textuel caché qui se trouve dans le cache de votre navigateur (Firefox, Chrome, etc.). Il peut comprendre des renseignements de diverses natures sur vous, selon notamment les sites que vous avez visités et selon les renseignements que vous donnez vous-même volontairement ou à

vosre insu (âge, sexe, adresse postale, préférences en matière de shopping, etc.). Tout site visité dépose des cookies dans votre machine.

➤ Notez que votre IP permet aussi de vous pister.

Les moteurs de recherche

Qu'est-ce que c'est ?

Bases de données alimentées automatiquement
par des robots
pour collecter et indexer
les pages web, les sites, les objets contenus
dans une page

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

« Certains sites web offrent un moteur de recherche comme principale fonctionnalité; on appelle alors « moteur de recherche » le site lui-même. Ce sont des instruments de recherche sur le web sans intervention humaine, ce qui les distingue des annuaires. Ils sont basés sur des « robots », encore appelés « bots », « spiders », « crawlers » ou « agents », qui parcourent automatiquement les sites à intervalles réguliers afin de découvrir de nouvelles adresses (URL). Ils suivent les liens hypertexte qui relient les pages les unes aux autres, les uns après les autres. Chaque page identifiée est alors indexée dans une base de données, accessible ensuite par les internautes à partir de mots-clés. » (Source : Wikipédia)

Les robots parcourant le web alimentent une base de données dans laquelle le moteur de recherche va piocher lors de chacune de vos recherches.

Les moteurs de recherche

Comment ça marche ?



Moissonnage (ou crawling) en continu pour récupérer :

- Le texte intégral d'un site
- Le contenu de l'entête du code source du site

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

Le robot d'un moteur de recherche va bien sûr ratisser toute la partie visible par un utilisateur humain dans une page web : texte, image, vidéo, animation, son ou tout autre objet qu'elle contient.

Mais en tant que machine, il va aussi lire et prendre en compte le **code source** (<https://fr.wikihow.com/voir-le-code-source>) de la page ainsi que les **métadonnées (1)** qui lui sont attachées.

Le **code source** d'une page internet est sa version en langage informatique qui comprend toutes ses instructions de mise en forme et **plus encore**. Chacune des instructions est signalée par une balise entre crochets < >

Et plus encore, car dans la zone **<head>** (en-tête), vous trouverez par exemple des mots-clés sur lesquels le moteur de recherche va se baser (comme la balise meta name="keywords", pour les mots-clés, ou un titre ou résumé que le moteur de recherche présentera sur sa page de résultats avec la balise meta name="description"). Dans Mozilla Firefox, exercez-vous sur la page <https://www.vidal.fr/parapharmacie/utilisation/bon-usage-phytotherapie-plantes/aromatherapie.html> : faites un clic droit, sélectionner « Code source de la page » et repérez les zones dont nous venons de parler.

Point Pix: Dans la zone <head> peuvent apparaître les instructions « noindex » ou « nofollow » ; cela indique au robot du moteur de recherche qu'il doit passer son chemin : la page concernée n'apparaîtra jamais dans une liste de résultats. Cela explique en partie pourquoi certaines pages n'apparaissent pas dans Google (ou un autre moteur)! (2) Pour aller plus loin: <https://www.commentcamarche.net/faq/9060-les-balises-dans-la-partie-head>

- (1) Les métadonnées sont des données informatives sur un objet électronique (page web, document PDF, image, vidéo, son, etc.) lisibles par la machine ou accessibles en consultant les **propriétés** du document. Ainsi les propriétés d'un document PDF nous renseignent sur ses dates de création et de modification, et les propriétés d'une image peuvent nous renseigner sur son auteur, la focale utilisée et l'appareil, voire sur le lieu photographié (coordonnées GPS) : toutes ces informations sont des métadonnées.
- (2) Voir diapo 22 la notion de « web invisible ».

Les moteurs de recherche

Où cherchent-ils ?

- Dimension du web ?

2 milliards de sites début 2023

- Web invisible

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY

Il existerait 4.72 milliards de pages individuelles indexées en octobre 2023 et près de 2 milliards de sites déclarés à ce jour (1).

Le web visible : est constitué par les sites indexables (répertoriables) par les outils de recherche. 10% du web total.

Le web invisible : (ou web profond) constituent les sites qui ne sont pas indexables par les outils de recherche :

- les pages web à accès réservé (code d'accès, identifiant)
 - les bases de données professionnelles interrogeables uniquement par des moteurs de recherche internes (ex : moteur de recherche propre au catalogue de la bibliothèque)
 - les archives des sites de presse d'actualité ou la page de réservation de la SNCF (pages dynamiques, qui changent en fonction de vos recherche ou de l'actualité)
 - les pages de sites trop volumineux
 - les pages déclarées par le webmestre non accessibles aux moteurs de recherche (cf. diapo 21)
 - les pages orphelines
- > Ne pas confondre avec le darknet (un réseau qui propose des fonctions d'anonymisation)

Même si les moteurs de recherche fonctionnent de la même manière, ils ont chacun leurs propres algorithmes pour le classement des résultats, qu'ils ne divulguent pas. C'est pourquoi l'on n'obtient pas les mêmes résultats d'un moteur à l'autre pour un seul mot clé. Vous n'obtiendrez pas non plus les mêmes résultats en fonction des postes informatiques (IP différente, cookies différents, géolocalisation différente, navigateur différent, historique différente dans le cache du navigateur, etc.). Les résultats varieront aussi dans le temps (et dans certains cas en fonction du lieu) pour un même utilisateur, sur une même IP et sur un même poste !

Les principes communs à tous les navigateurs sont tout de même connus.

(1) <https://news.netcraft.com/archives/category/web-server-survey/>;
<https://www.worldwidewebsite.com/> ; ou
<https://www.internetlivestats.com/watch/websites/> (en direct)

L'exemple choisi est Google (1)

L'ordre de classement des résultats est l'élément le plus important dans la mesure où les internautes se contentent majoritairement de la lecture de la première page de résultats.

Le taux de clics décroît très vite avec le positionnement du résultat, le premier lien capte 33 %, le 2e = 17,6 %, le 3e = 11,4 %.

La 1ère page reçoit 92% de clics. (2)

Quelques critères de classement des résultats :

- Nouveauté du contenu du site web
 - Affichage des résultats en fonction de l'historique de consultation et des réseaux sociaux
 - Il se fait par rapport à la requête de l'internaute (ordre des termes dans la requête...)
 - Présence du mot dans le titre de la page ou dans l'URL position du mot dans le premier tiers de la page
 - Fréquence d'apparition du mot dans la page, ou proximité des termes recherchés
 - Mise en valeur du mot dans la page (police, caractère gras, liens...)
 - Par popularité des liens : (le PageRank inventé par Google en 1998, cf. : <https://fr.wikipedia.org/wiki/PageRank>)
 - Mesure d'audience ou indice de clics (pages populaires)
 - Positionnement payant des liens contextualisés
 - **Sites commerciaux (annonces) en tête de page**
 - Sites sécurisés (protocole https) -> voir : https://fr.wikipedia.org/wiki/HyperText_Transfer_Protocol_Secure
 - Le site s'adapte-t-il indifféremment aux ordinateurs, tablettes, smartphones ...?
- L'expérience montre malheureusement que les résultats sont souvent pollués en première position par des liens publicitaires, commerciaux ou tout simplement populaires (et la popularité n'a rien à voir avec la rigueur scientifique!)

(1) Pourquoi présenter Google plutôt qu'un autre ?

- Part de visites des moteurs de recherche en France : Chiffres juillet 2022 statcounter : google : 91,31%, bing : 4,5%, yahoo : 1,37, 0,82% Ecosia (montée en puissance des moteurs de recherche 'écologiques') ; Qwant : 0,66, DuckDuckGo : 0,59 (légère progression de Google et des moteurs protégeant les données comme Duckduckgo et Qwant).

Voir : <https://gs.statcounter.com/search-engine-market-share>

Et aussi : <http://www.planetoscope.com/Internet-/1474-recherches-sur-google.html>

- Google crawl 20 milliards de pages /jour et reçoit 80 000 requêtes chaque seconde, soit 6,9 milliards par jour (15 % des requêtes sont de nouvelles requêtes (500 millions par jour)) ! <https://www.blogdumoderateur.com/chiffres-google/>
- 200 algorithmes développés mais inconnus, avec de nombreuses modifications par an. <https://www.webrankinfo.com/dossiers/conseils/algos-google>
- Equipe considérable, moteur en perpétuelle évolution

En outre, sa part est écrasante sur le marché :

<https://www.leptidigital.fr/webmarketing/seo/parts-marche-moteurs-recherche-france-monde-11049/>

(2) <https://www.blogdumoderateur.com/etude-taux-de-clics-et-les-performances-des-liens-selon-la-position-dans-les-serps/#:~:text=Le%20CTR%20de%20la%201%C3%A8re%20position%20est%20de%2031%2C7%25&text=V%20ici%20comment%20sont%20partag%C3%A9s%20en,donec%20encore%20une%20importance%20%C3%A9norme.>



Google a fait des petits ! (1)

- Google images = propose depuis 2001, affichage sous forme de vignettes avec taille, poids format et source mais permet aussi la recherche inversée (2)
- Google actualités = va chercher les informations en temps réel chez les grands médias (500 sources francophones).
- Google books = accès à des livres numérisés (numérisation depuis 2004), surtout des Etats-Unis
- Googlemaps
- Youtube
- Google scholar = depuis 2004, porte sur des contenus scientifiques : articles scientifiques approuvés ou non par des comités de lecture (peer-reviewed), des thèses de type universitaires, citations ou encore des livres scientifiques. Permet d'obtenir le lien vers le texte intégral quand il est disponible. **Cependant, cela reste du Google : si la majorité des résultats est sérieuse, il convient de rester vigilant (on peut parfois trouver du contenu non scientifique qui veut se faire passer pour tel, ou encore de vrais articles scientifiques mais dans une version non encore définitive ou tronquée, etc., etc. : apparaître dans Google Scholar ne garantit pas la qualité du contenu).** (3)

(1) Ce site fournit une liste un peu plus complète pour 2023 : <https://larevuetech.fr/services-google-2023/>

(2) **Point Pix** : Internet et les réseaux sociaux sont inondés d'images présentées comme preuves d'une information souvent prétendument sensationnelle. La recherche d'image inversée dans Google Image ou dans Bing! permettent de retrouver une photo dans son contexte original ou de démasquer les photos truquées. Le site Tineye <https://tineye.com/> est aussi intéressant à ce sujet.

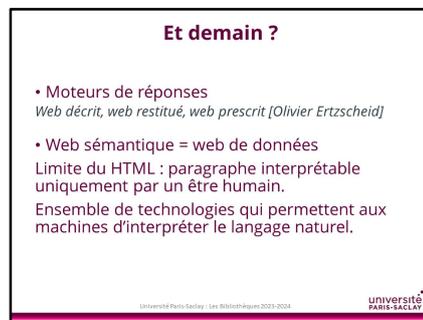
-> Pour en savoir plus sur la recherche d'image inversée : <https://www.geekmaispasque.com/2020/08/faire-recherche-image-inversee/>

-> Notez que consulter les métadonnées d'une photo dans le menu Propriétés (auteur, date de prise de vue, coordonnées GPS...) peut être **parfois** révélateur : téléchargez l'image sur votre ordinateur, faites un clic droit sur le fichier fermé, sélectionnez « Propriétés » et consultez l'onglet « détails »).

-> Ce bref tutoriel vidéo ne manque pas d'intérêt : <https://www.20minutes.fr/high-tech/2246663-20180402-video-fact-checking-trois-conseils-identifier-image-internet>

(3) Pour des articles plus fiables et évalués selon leur niveau de preuve, **TRIP** est un moteur

intéressant dans le domaine de la santé (on gagne ici en qualité par rapport à Google Scholar) :
<https://www.tripdatabase.com/>



- **Aujourd'hui et hier : les moteurs de réponse**

Plusieurs étapes ont mené à la recherche d'informations sur internet telle que nous la pratiquons sur les moteurs de recherche :

- Au début, « le web décrit » : comme par exemple dans un annuaire de sites ;
- Ensuite, « le web restitué » : les pages de résultats mises en forme par les moteurs de recherche pour être lisible par l'utilisateur (=vous) ;
- Et aujourd'hui : « le web prescrit » : le moteur de recherche « essaie » de deviner vos désirs, ce que vous voulez « réellement ».
- Une quatrième étape est déjà là, « le web prédictif ». Le moteur de recherche, grâce aux données disséminées sur vous dans le **Big data** (1), grâce à vos cookies, grâce à votre historique de recherche, grâce à votre IP, etc. va « tenter » de « prédire » ce qui pourra vous plaire, avant même que vous ne le sachiez, voire de faire émerger des envies ou des besoins que vous n'auriez pas eus autrement (annonces et publicités dans Google ou tout autre moteur commercial) ; vous voyez comme le système est ambigu : d'un côté cela semble pour votre bien (personnalisation optimale du service) d'un autre, on pourrait être tenté de penser que cela est de la manipulation d'opinion ou une incitation à consommer sans besoin réel (2).

Exemple de résultat prescrit : vous cherchez un chapeau melon sur Ebay et le lendemain, lorsque vous allez sur Youtube, on vous propose une vidéo sur l'actrice Emma Peel de la série « Chapeau melon et bottes de cuir » et une autre ensuite sur les comiques Laurel & Hardy (qui portent des chapeaux melons) au milieu de votre sélection musicale habituelle qui n'a strictement rien à voir ; bref, Youtube a « cherché » à satisfaire votre goût prononcé pour ce type de couvre-chef. Notez qu'Ebay et Youtube sont deux sites distincts : ce sont vos cookies et votre historique qui vous ont « trahis » (3) (4) car le cache de votre navigateur est lisible par tout site visité !

On s'achemine vers le **web sémantique**, basé sur une intelligence artificielle (5) censée comprendre les modes de pensée et d'expression humaine. (Voir le traducteur DeepL, ou le moteur de recherche Wolframalpha (<https://www.wolframalpha.com/>)).

(1) <https://www.futura-sciences.com/tech/definitions/informatique-big-data-15028/>

(2) https://fr.wikipedia.org/wiki/Bulle_de_filtres

(3) **Pensez aux fonctionnalités de navigation privée de votre navigateur, nettoyez de temps en temps le cache** de votre navigateur, et lorsqu'un site vous propose d'accepter ses cookies, réfléchissez-y à deux fois avant de cliquer sur « tout accepter » (lorsque c'est faisable, optez pour un paramétrage personnalisé des cookies). C'est d'ailleurs depuis le RGPD (diapo 19) qu'il nous est proposé d'accepter, de refuser ou de sélectionner les cookies d'un site. Des paramètres avancés de votre navigateur permettent aussi de refuser les cookies. Pour cela, voir : <https://www.cnil.fr/fr/les-conseils-de-la-cnil-pour-maitriser-votre-navigateur>

(4) Pour aller plus loin : <https://www.cnil.fr/fr/publicite-ciblee-en-ligne-quels-enjeux-pour-la-protection-des-donnees-personnelles>

(5) Ce support ne prend malheureusement pas en compte ChatGPT et ses amis ou cousins plus ou moins proches, le formateur n'a pas assez de recul : les changements sont allés trop vite!

Bibliothèque universitaire



Si vous avez des questions :

bib.univ@universite-paris-saclay.fr

Université Paris-Saclay : Les Bibliothèques 2023-2024

université
PARIS-SACLAY