

Université Sétif 02 - Mohamed lamine debaghine

Faculté des Lettres et des Langues

Département de langue française

Année universitaire 2023/2024

Niveau: Master 01

Spécialité: Didactique des langues étrangères

# Recherche documentaire sur web, bibliographie

- Réalisé par: Elaggoune Zakarya
- @ mail: zelaggoune@gmail.com
- <https://cte.univ-setif2.dz/moodle/>

# Objectifs du module

**Parite 01: Recherche  
documentaire sur Web**

**Parite 02: Gestion de  
références bibliographiques**

# Objectifs du module - partie 01 : Recherche documentaire sur Web

- Cette première partie vise principalement à permettre à l'apprenant de :
  - 1) Identifier les sources pertinentes dans un domaine donné, en particulier celles en accès libre sur Internet.
  - 2) Optimiser la recherche documentaire en tirant la meilleure partie des moteurs de recherche.

# Objectifs du module - partie 01: Recherche documentaire sur Web

Les moteurs de recherche



# Objectifs du module - partie 02: Gestion de références bibliographiques

- A l'issue de cette deuxième partie du module, l'apprenant doit être capable de:
  - 1) Construire et alimenter une bibliographie ordonnée.
  - 2) Citer convenablement ses sources bibliographiques pour éviter le plagiat.

# Objectifs du module - partie 02: Gestion de références bibliographiques

*Éditeur de texte*



*Logiciels de gestion  
de références  
bibliographiques*



# Plan du module

- (1) L'Internet et la recherche documentaire**
- (2) Effectuer une recherche sur Google**
- (3) Effectuer une recherche sur Google scholar**
- (4) Comment faire une bibliographie dans un document avec Microsoft Word**
- (5) Les logiciels de gestion de références bibliographiques**

# **1) L'Internet et la recherche documentaire**

# Qu'est-ce que la recherche documentaire ?

- **La recherche documentaire** est une étape de travail à réaliser avant de se lancer dans une étude empirique.
- Elle permet de **collecter** des données informatives grâce à l'étude de documents issus de sources fiables, de documents officiels ou universitaires.
- Les informations récoltées seront utiles pour développer des connaissances sur le sujet étudié.

# Objectifs de la recherche documentaire

- *La finalité d'une bonne méthodologie de recherche est de faciliter la production d'un travail universitaire alliant richesse documentaire et rigueur scientifique.*

# Qu'est-ce que la recherche documentaire ?

- **Les documents collectés dans la recherche documentaire peuvent être de différents types :**
  - Monographies (livres),
  - Mémoires,
  - Thèses,
  - Actes de colloques,
  - Rapports,
  - Articles de périodiques,
  - Archives (d'auteurs ou d'institutions),
  - Documents audiovisuels, Images,
  - Données statistiques,
  - Arrêtés et autres textes juridiques (circulaires, lois, décrets)...

# Dimensions de la recherche documentaire

## Méthodologique

- La préparation de la recherche
- Chercher les références et vérifier leur véracité
  - Éviter le plagiat

## Technique

- L'utilisation des moteurs de recherches
- Citer correctement les sources
  - Présenter une bibliographie

**Dans ce cours, nous allons nous concentrer beaucoup plus sur le côté technique**

# Les documents électroniques

- Les documents électroniques également appelés documents numériques sont consultables à partir d'un ordinateur.
- **Les sites web, les bases de données et les revues électroniques** sont des documents électroniques qu'on utilise, au même titre que les documents papier tels que les livres ou les périodiques.

# Les documents électroniques

- **Les sites web:**
  - **Les sites web sont des document électroniques organisés en pages web reliées entre eux par des liens hypertextes. La page d'accueil présente le site et propose les premiers liens hypertextes.**

# Les documents électroniques

- **Les bases de données:**

- **Les bases de données sont des documents électronique qui nous permettent d'accéder simultanément à des références mais aussi au texte intégral de plusieurs revues.**

# Les documents électroniques

- **Les périodiques électroniques:**
- **Les périodiques en version papier existent souvent sous une «version électronique» Certains périodiques électroniques peuvent n'exister que sous cette forme, sans jamais avoir été édités sur un support papier.**

# Les documents électroniques

- **Question :** Comment pouvons nous trouver les documents électroniques pertinents à nos besoins ?
- **Réponse:** grace à l'Internet, le Web et les moteurs de recherches

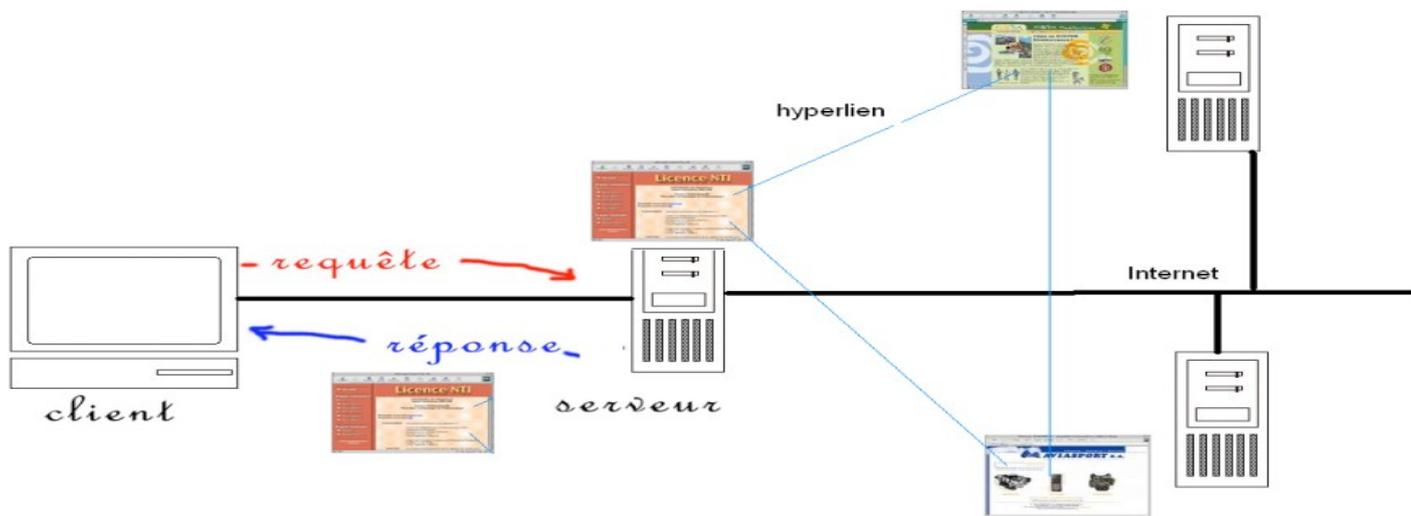
# L'Internet et l'architecture Client/Serveur

- L'Internet est un ensemble de réseaux de toutes tailles composés d'ordinateurs hétérogènes et interconnectés en utilisant un langage commun appelé TCP/IP (Transmission Control Protocol/Internet Protocol)



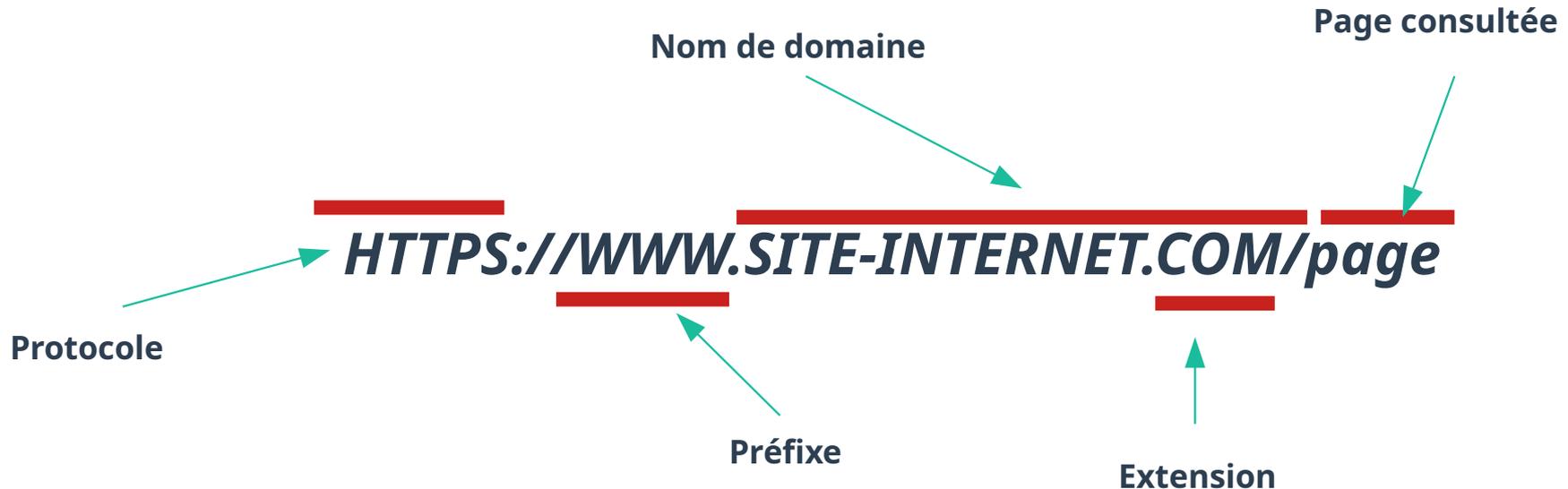
# Qu'est ce que le Web alors???

- **Définition simple :** « *Le Web est un réseau de documents reliés entre eux par des liens hypertextes. Chaque page est située physiquement sur un noeud (serveur) du réseau mondial Internet* »



# Requête : Comprendre l'URL d'un site

- Une adresse URL (Uniform Resource Locator) est composé de:



# Requête : Comprendre l'URL d'un site

- Exemple : `https://cte.univ-setif2.dz/moodle/`
- Protocole: `https`
- Domaine: `univ-setif2.dz`
- Sous-domaine: `cte.univ-setif2.dz`
- Page consultée: `moodle`

# Que trouve-t-on sur le Web ?

- **Sur l'Internet, on peut trouver tout type de document :**
  - du texte,
  - des images simples, animées ou en 3D
  - du son
  - des vidéos
  - des programmes
  - ...
- **Les navigateurs récents intègrent désormais de nombreux "plug-ins" permettant de visualiser certains documents.**

# Caractéristiques de l'information

- **L'information sur Internet est:**
  - destinée aux humains
  - hétérogène (contenu et format)
  - instable et éphémère (liens morts, changement d'adresse...)
  - Multilingue
  - renouvelée continuellement
  - gratuite ou payante
  - non structurée (pas un format uniforme)

# Résumons la situation

- **On a des machines reliées entre elles :**
  - chacune est uniquement identifiable par un nom (adresse URL)
  - chacune héberge des documents contenant de l'information destinée à être public
  - Chaque document est destiné à un public spécifique
  - Chaque document publié n'est pas forcément fiable

# La problématique

- **Nous avons aujourd'hui des milliards de pages, et la question qui se pose est la suivante :**  
comment pouvons-nous trouver **les documents pertinents** dans cette importante masse d'informations ? Et comment vérifier la **véracité** et la **fiabilité** de ces documents ?

# La solution

- **Trouver un moyen d'organiser le contenu du Web**
- **Automatiser le processus de localisation des pages Web**

# Modes d'accès possible à l'information

- 1) La recherche par navigation en utilisant un répertoire thématique**
  - Classification thématique (figée) manuelle, ressources limitées (intervention humaine irréaliste face à la masse d'information)
- 2) La recherche par mot-clé en utilisant un moteur de recherche**
  - Indexation automatique, plus exhaustif mais sans classification (totalement automatique)
- 3) La recherche de type spécifique d'information (document)**
  - Moteurs de recherche spécialisés (son, vidéo, etc.)

# La recherche par navigation en utilisant un répertoire thématique

- **Les répertoires sont des ressources classées par sujet dans une base de données hiérarchisée du plus général au plus spécifique**
- **Ce type de classement permet à l'utilisateur de préciser progressivement sa recherche en passant successivement d'une catégorie à une sous-catégorie.**
- **Classification faite par des Humains**

## Our business is publishing

With more than 2,900 journals and 300,000 books, Springer offers many opportunities for authors, customers and partners.

### Explore our subjects



**Read** over ten million scientific documents on [»SpringerLink](#).

**Buy** more than 300.000 different books in our [»SpringerLink Shop](#) ↗

- » [Astronomy](#)
- » [Behavioral Sciences](#)
- » [Biomedical Sciences](#)
- » [Business & Management](#)
- » [Chemistry](#)
- » [Climate](#)
- » [Computer Science](#)
- » [Earth Sciences](#)
- » [Economics](#)
- » [Education & Language](#)
- » [Energy](#)
- » [Engineering](#)
- » [Environmental Sciences](#)
- » [Food Science & Nutrition](#)
- » [General Interest](#)
- » [Geography](#)
- » [Law](#)
- » [Life Sciences](#)
- » [Materials](#)
- » [Mathematics](#)
- » [Medicine](#)
- » [Philosophy](#)
- » [Physics](#)
- » [Public Health](#)
- » [Social Sciences](#)
- » [Statistics](#)
- » [Water](#)

# Moteurs de recherche

- **Un moteur de recherche est l'outil à utiliser quand on cherche une information précise qui peut être située sur une page particulière d'un site**
- **Un moteur de recherche est divisé en quatre parties:**
  - un programme appelé robot (ou bot) qui scanne les pages pour récupérer leurs données (les Web crawler)
  - une base de données pour stocker ces données (à de rares exceptions près, exclusivement des textes)
  - une structure qui organise ces pages (indexation)
  - une interface (une page Web) permettant de faire des recherches d'informations dans la base de données.

# Moteurs de recherche

- **Pour utiliser un moteur de recherche, il suffit de :**
  - connaître l'adresse du moteur de recherche
  - savoir ce qu'on cherche !!!!
  - savoir comment formuler la requête (l'équation de recherche)

# Moteurs de recherche

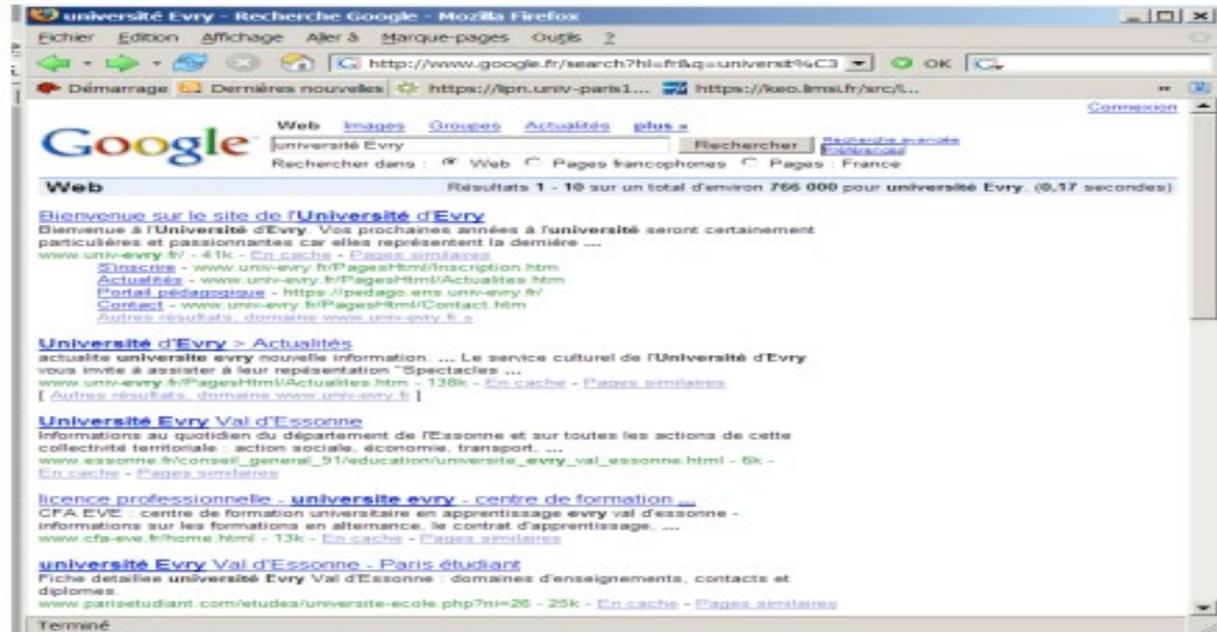
- **La recherche s'effectue sur les mots contenus dans les ressources elles-mêmes**
- Cette recherche affiche les sites correspondant aux termes demandés par le navigateur suivant un certain ordre qui dépend de divers paramètres (importance du site, nom, nombre de fois que le mot est repris dans la page, ...) et dans certains cas les sites qui payent (liens sponsorisés)

# Recherche par mots-clés

- **Pour mieux chercher il faut exprimer les mots-clés en levant le maximum d'ambiguïté, et cela avec les fonctionnalités des moteurs de recherche :**
  - ➔ **Possibilité de recherche avec des opérateurs booléens (AND, OR, NOT)**
  - ➔ **Spécifier le type de document**
  - ➔ **Position des mots dans la page (titre, corp,...)**
  - ➔ **Possibilité de recherche par groupes de mots**
    - Exemple : "traitement de l'information"
  - ➔ **Possibilité de recherche par troncation**
    - Exemple : Etudiant\* --> Etudiant, Etudiante, Etudiants, Etudiantes

# Exemples de moteur de recherche

- Google
- Lycos
- Altavista
- Yahoo



- Remarque : plusieurs portails permettent la recherche par mots-clés ainsi que la navigation thématique, exemple : Yahoo, Voila, etc.

# Moteurs de recherche spécialisés

- **Recherche dans le domaine scientifique:**
  - <https://scholar.google.com/>
- **Recherche d'images et de sons:**
  - Image Search, d'autres moteurs de recherche « généraux » proposent ces fonctionnalités (eg. Google)
- **Recherches de logiciels:**
  - [shareware.com](http://shareware.com)
- ...

# Pertinence des documents retournés -

- **Pertinence des documents retournés -**

- on parle de “**bruit**” lorsque des réponses non-pertinentes sont proposées par le moteur de recherche. Ces réponses sont mêlées à des réponses pertinentes, qui risquent alors de ne pas être vues par l'utilisateur.
- Les causes du bruit peuvent être multiples :
  - l'utilisateur se sert de termes de recherche possédant plus d'un sens, par exemple, la couleur « orange » et le fruit « orange » (problème de polysémie);
  - l'utilisateur n'utilise pas assez de termes dans sa requête ;
  - l'utilisateur formule une requête comprenant trop de termes coordonnés par l'opérateur OU (l'opérateur booléen OR).

# Pertinence des documents retournés -

- **Pertinence des documents retournés -**
  - ➔ on parle de “**silence**” lorsque des réponses pertinentes ne sont pas proposées par le moteur de recherche
  - Les causes du silence peuvent être multiples :
    - un même concept est représenté par plus d'un terme de recherche, par exemple, « football » et « soccer » (problème de synonymie);
    - l'utilisateur formule une requête comprenant trop de termes coordonnés par l'opérateur ET (l'opérateur booléen AND) ;
    - la requête comportent des fautes d'orthographe.

# Critères d'évaluation des méthodes d'accès

- **Temps de réponse**
- **Interface**
  - Conviviale (simple à utiliser) ?
  - Intuitive ?
  - Aide en ligne ?

# Le Web invisible

- Généralement les fichiers qui ne sont pas au format HTML
- Les pages Web demandant un login ou restreintes à un domaine particulier
- Les pages dont le contenu est dynamique (créées en réponse à la demande d'un utilisateur par exemple)
- Les sites intranet
- Les pages qui ne sont pas référencées (aucun lien ne pointe vers cette page)
- Les ressources ne provenant pas du Web (site FTP, gopher...)
- ...

# Annexe

- La signification des messages d'erreurs
  - 400 Bad Request
  - 401 Unauthorized
  - 402 Payment required
  - 403 Forbidden
  - 404 Not Found
  - 500 Internal Server Error

# Je m'interroge !



- Lors de la réalisation et de la rédaction d'un mémoire de fin d'études, comment procédez-vous pour mener une recherche documentaire correcte ?
- Comment vérifier et assurer la véracité des documents trouvés ?