

Les moteurs de recherche

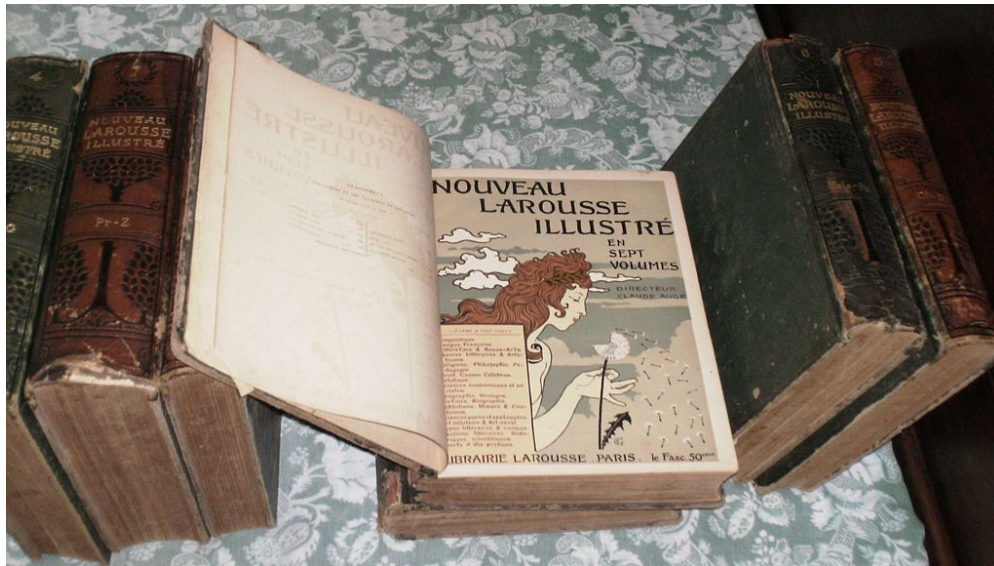
Questions préliminaires

- Quel est le moteur de recherche le plus utilisé actuellement ?
- Pouvez-vous citer d'autres moteurs de recherche ?

Qu'est-ce qu'un moteur de recherche ?

Qu'est ce qu'un moteur de recherche ?

- Un moteur de recherche est une application qui permet de trouver des informations dans des pages dont on ne connaît pas l'adresse (l'URL), voire dont on ignore l'existence.
- Véritable révolution de l'accès à l'information !



Le Nouveau Larousse illustré (1897-1904).
Crédit : André Koehne, Domaine public, via Wikimedia Commons.



Encyclopédie spécialisée sur le monde gréco-romain
Crédits : Christoph Alexander Martsch, licence CC BY-SA 3.0, via
Wikimedia Commons

Fonctionnement d'un moteur de recherche

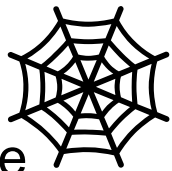
Fonctionnement d'un moteur de recherche

- Concrètement on lui pose une question et il nous répond quelque chose du genre :
« sur les 100 milliards de documents que j'ai, il y en a 5 720 552 qui correspondent à la question et voici les 10 plus intéressants ! »
et le tout en un dixième de seconde !
- Le travail se décompose en 3 grandes étapes :
 - ◆ *« sur les 100 milliards de documents que j'ai »*
→ étape 1 : collecte des pages existantes
 - ◆ *« il y en a 5 720 552 qui correspondent à la question »* et *« les 10 plus intéressants »*
→ étape 2 : analyse du contenu des pages et classement des pages selon leur pertinence
 - ◆ *« on lui pose une question et il nous répond »*
→ étape 3 : construction de la liste ordonnée des pages en réponse à la question

Fonctionnement d'un moteur de recherche

- **Étape 1** : Exploration du Web

Des robots, qui sont en fait des programmes informatiques, parcourent le Web en permanence pour collecter toutes les pages existantes.



- **Étape 2** : Indexation et classement des pages



- ◆ Les pages collectées sont analysées et indexées sur les mots qu'elles contiennent. L'index associe à chaque mot clé les URL des pages où il apparaît.

informatique <https://fr.wikipedia.org/wiki/Informatique>
<https://www.ldlc.com/informatique/cint3063/>
<https://www.onisep.fr/Decouvrir-les-metiers/des-metiers-par-secteur/Informatique-et-reseaux/Les-metiers-et-l-emploi-dans-l-informatique-et-les-reseaux>

...

web https://fr.wikipedia.org/wiki/World_Wide_Web
https://fr.wikipedia.org/wiki/Site_web

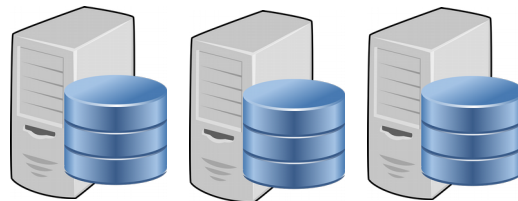
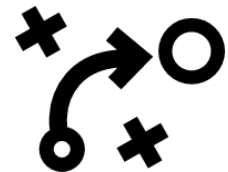
...

```
abs, 434
accumulateur, 30, 31
add, 190
additionneur (circuit), 281
affectation, 7
agrégation, 215
algèbre de Boole, 279
algorithme
  de tri, 143
  des plus proches voisins, 175
  glouton, 161
aliasing, 81
allumettes (jeu des), 101
ALU, 288
and, 48
Android, 310
année bissextile, 55
anniversaire, 189
appel d'une fonction, 58
append, 77, 217
Apple, 310
```

Index d'un livre

Fonctionnement d'un moteur de recherche

- **Étape 2** : Indexation et classement des pages (suite)
 - ◆ Les pages sont ensuite classées de la manière la plus pertinente possible. Ce classement dépend notamment de :
 - ◆ La **pertinence** de la page (par exemple, l'importance du mot cherché dans cette page : c'est mieux s'il se trouve dans un titre que dans un paragraphe)
 - ◆ La **popularité** de la page
 - ◆ Ce sont des **algorithmes** qui calculent ce classement
→ concurrence acharnée entre moteurs de recherche !
- Tout le travail et les calculs des étapes 1 et 2 est fait à l'avance :
 - ◆ permet de répondre en un temps très court (étape 3)
 - ◆ nécessite de posséder des serveurs très puissants et avec énormément de mémoire



Fonctionnement d'un moteur de recherche

- **Étape 3** : Réponse aux requêtes des internautes
 - ◆ Le moteur de recherche récupère les mots clés saisis par l'internaute et fabrique une liste ordonnée d'adresses
 - ◆ Les pages proposées à l'internaute sont triées par qualité et pertinence selon les algorithmes de chaque moteur de recherche

The screenshot shows the Qwant search engine interface. The search bar contains the word 'informatique'. Below the search bar, there are navigation tabs for 'Web', 'Actualités', 'Images', 'Vidéos', 'Shopping', 'Maps', and 'Musique'. The results are filtered for 'France' and 'Toute période'. The first result is from 'Rue du Commerce' with the title 'Rue du Commerce® Site Officiel - Le meilleur du High-Tech'. The second result is from 'Boulanger' with the title 'Boulanger® - Electroménager et Multimédia'. The third result is from 'darty.com' with the title 'Informatique - Livraison gratuite, retrait 1h en magasin'. The fourth result is from 'fr.wikipedia.org' with the title 'Informatique'.

The screenshot shows the Microsoft Bing search engine interface. The search bar contains the word 'informatique'. Below the search bar, there are navigation tabs for 'TOUT', 'IMAGES', 'VIDÉOS', 'CARTES', 'ACTUALITÉS', 'SHOPPING', and 'À PROPOS DES RÉSULTATS DE RECHERCHE'. The results are filtered for 'Date' and 'Résultats à proximité de Chaumont-d'Anjou, Pays de la Loire'. The first result is from 'commentcamarche.net' with the title 'Informatique - Définition - Comment Ça Marche'. The second result is from 'ldlc.com' with the title 'Informatique - Achat Informatique - LDLC'. Below the results, there is a carousel of product categories: 'Image & son', 'Téléphonie & auto', 'Jeux & Geek Zone', and 'Objets connectés'. The third result is from 'ruecommerce.fr' with the title 'Informatique : PC - Composants - Périphériques | RueDuCommerce'.

Le saviez-vous ?

- Les premiers moteurs de recherche ont été créés dans les années 1994-1995



Crédit : Tim42, Domaine public via Wikimedia Commons



Crédit : AltaVista Webseite, Domaine public

- Les 3 premiers résultats d'un moteur de recherche remportent plus de % des clics
- Le moteur de recherche Google indexe actuellement près de de pages
- Il existe des pays dans lesquels Google n'est pas le moteur le plus utilisé ! En et en par exemple.

Crédits

- **7 ht** [uxwing](#)
7 m [uxwing](#)
- **8 ht** [uxwing](#)
8 bas [OpenClipart-Vectors](#), via Pixabay
- **10 g** [Tim42](#) Domaine public via Wikimedia Commons
10 d [AltaVista Webseite](#), Domaine public, via Wikimedia Commons

Références

- Gérard Berry, « Pourquoi l'informatique met le monde à l'envers », conférence donnée à l'université de Strasbourg : [en vidéo](#).
- Manuel SNT, collection 3.0, BORDAS.
- Vincent Brossas, LEPTIDIGITAL, « Les parts de marché 2022 des moteurs de recherche en France et dans le Monde ! », source : [leptidigital.fr](#).