

WEB : RECHERCHE DE PAGE WEB

LE FONCTIONNEMENT D'UN MOTEUR DE RECHERCHE

Les moteurs de recherche permettent de trouver des informations dans des pages dont on ne connaît pas l'adresse, voire dont on ignore l'existence. Le plus utilisé étant Google (en 2018, 90% des internautes l'ont utilisé dans le monde).

Exercice : Effectuer la recherche « Apollo 11 » sur les quatre moteurs de recherche proposés et compléter le tableau suivant avec vos résultats de recherche

| Moteurs de recherche | Position éventuelle de publicités | Position du premier lien vers wikipédia | Position et nom du premier lien vers un média culturel | Position du premier lien vers le film Apollo 11 |
|----------------------|-----------------------------------|---|--|---|
| Google | | | | |
| Qwant | | | | |
| Ecosia / bing | | | | |
| Million short | | | | |

En déduire celui que vous conseilleriez. Argumenter.

Etapas de fonctionnement

Les moteurs de recherche fonctionnent en quatre temps.

Deux avant la demande l'internaute

- Collecte des informations par des robots (crawlers ou spiders)
(Analyse du code html, vérification des liens internes et externes)
- Indexation et référencement des pages web récoltées (Base de données)

Deux après la demande de l'internaute

- Recherche de pertinence des pages (algorithme)
- Affichage des pages proposées



<https://www.youtube.com/watch?v=Y8l4hKNQOEY>

Les **robots** sont des processus informatiques qui parcourent une page web, y récupèrent les liens hypertextes afin d'aller visiter les pages liées ultérieurement. En suivant ces liens un par un, ils vont parcourir pratiquement toute la toile (d'où le nom de **spider** ou **crawler**)

Le ranking

Le système de ranking est la méthode permettant de classer les pages web. Les algorithmes de chaque moteur de recherche sont considérés comme secret industriels et jalousement gardés.

Néanmoins on peut distinguer 3 types de tri

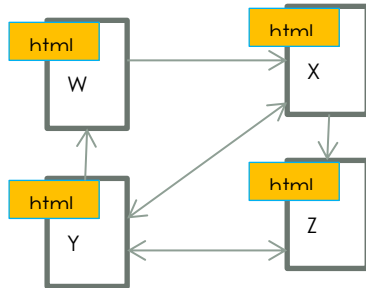
- Le tri par pertinence qui évalue le nombre de fois où un mot ou un document apparaît dans la page web. L'utilisation de balises HTML spécifiques (<h1> ...) permettent de mettre une valeur des mots clés

Fast Facts: **Apollo 11**
 The goal of **Apollo 11** was, as John F. Kennedy had pronounced less than a decade earlier, to land astronauts on the Moon and return them safely to Earth.

The mission **Apollo 11** conducted the first crewed landing on the Moon, deployed instruments, took photographs, collected samples and returned the crew safely back home.

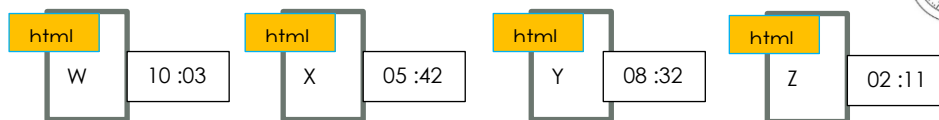
Dans le petit texte ci-contre, nous dénombrons occurrences du mot clé Apollo

- Le tri par popularité qui évalue le nombre de liens qui pointent vers la page web (PageRank)



Dans l'exemple ci-contre, un robot crawler aura plus de chance de se retrouver sur la page

- Le tri par audience qui évalue le temps resté par les internautes sur la page web



Sachant que 80% des internautes ne consultent que les trois premiers résultats d'une recherche, des sociétés se sont spécialisées en référencement afin qu'une page web soit présente parmi les premières propositions de recherche. Parmi les méthodes utilisées, il existe le spamming : En écrivant des mots clés en blanc sur fond blanc, ils agissent sur le tri par pertinence.

Soyez critique et osez explorer les pages web mal référencées. C'est ce que propose million short qui se présente plus comme un moteur de découverte qu'un moteur de recherche.