

1. Présentation générale

Les moteurs de recherche sont indispensables au bon fonctionnement du Web. Ils permettent de trouver des informations dans des pages dont on ne connaît pas l'adresse, voire dont on ignore l'existence.



Parlons un peu de Qwant

Qwant est un **moteur de recherche créé en France** et lancé en version bêta le 16 février 2013. **Contrairement à Google**, qui exploite vos données personnelles, lorsque vous effectuez une recherche **sur Qwant, aucun cookie de traçage n'est déposé sur votre navigateur**. La recherche n'est pas stockée sur le serveur, ni associée à votre profil pour être réutilisée plus tard.

2. Comment fonctionne un moteur de recherche ?

2.a) Collecter les données

A chaque moment, **des robots d'exploration "crawlers" ou "spiders" explorent le web à la recherche de page**. C'est à dire qu'ils visitent les pages du web et vont suivre tous les liens qu'ils trouvent afin de découvrir de nouvelles pages (que ce soit au sein du même site ou bien si les liens renvoient vers d'autres domaines, des nouveaux site).

Les robots sont très friands de contenu frais et nouveau, plus vous en créez, plus les robots viendront "crawler" votre site régulièrement.

2.b) L'indexation

Une fois les données collectées, les moteurs de recherche vont s'atteler à les stocker dans ses datacenters, mais surtout à les organiser.

C'est le travail d'indexation qui commence : les moteurs vont étudier en détail toutes les pages qu'ils ont pu recevoir (le code, les images, les liens, les mots clés ...) et vont classer toutes ces informations en deux index différents :

- **index principal** : index dans lequel le moteur de recherche va mettre toutes les données récoltées durant le crawl.
- **index inversé** : va contenir tous les mots-clés potentiels que l'on peut associer aux URL du web.

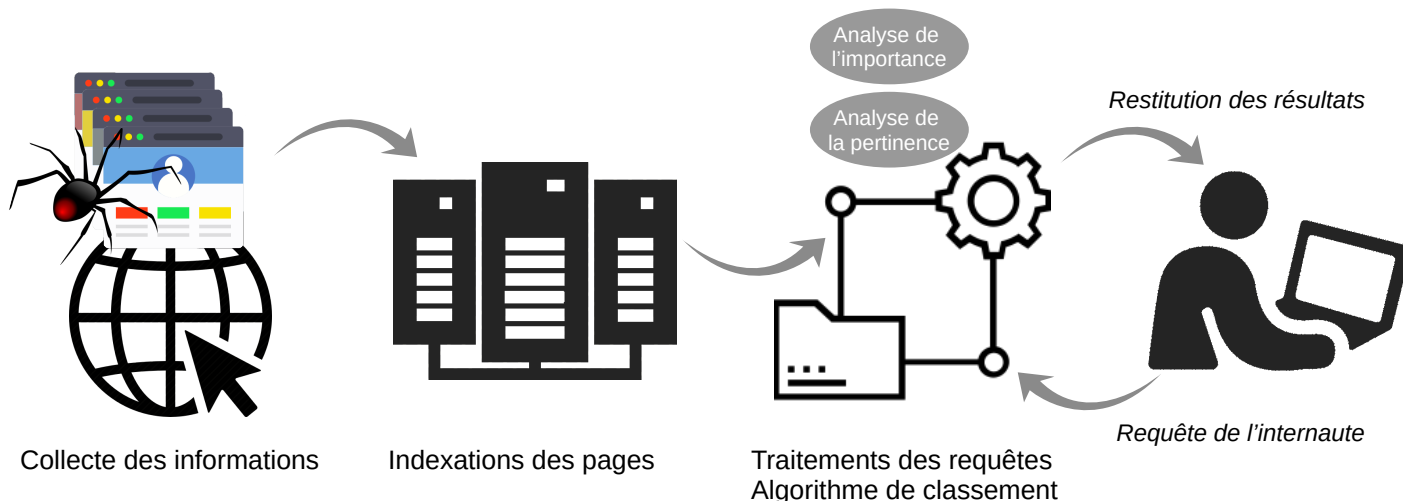
Cet index va permettre aux moteurs de recherche de connaître le nombre de fois qu'un mot-clé apparaît dans une page, en comparaison à d'autres pages et bien entendu, cela aura une influence sur classement de la page lorsqu'un internaute tapera ce mot clé. Attention, faire du "bourrage de mots clés (Keyword Stuffing)" est pénalisé par les moteurs de recherche, notamment Google qui utilise des centaines d'autres critères.

2.c) Traitement des requêtes et Ranking

Après avoir collectées et indexées les données des sites, les moteurs de recherche vont chercher à proposer aux internautes les pages qui leur paraissent les plus pertinentes par rapport aux requêtes de ces derniers.

Pour cela, les moteurs de recherche font du **référencement naturel** : les données (donc les pages web) sont classées selon des centaines de critères que l'on peut regrouper de la manière suivante :

- **Pertinence** : cela va concerner les mots clés présent dans la page, leur densité, leur positionnement, le poids de ces mots clés dans la base de données, la proximité entre la requête des internautes et les mots clés présents sur la page...
- **Popularité** : correspond au Netlinking, c'est à dire le nombre de lien sur un site qui pointe vers un autre site. En bref, la popularité sera définie par le nombre de gens qui parle d'un site et la manière dont il parle de ce site.
- **Audience** : cela correspond à la qualité du trafic qui vient sur une page. Plus les internautes restent sur un site, cliquent sur des pages, reviennent dessus, plus ce site va être considéré comme pertinent et sera mieux référencé.



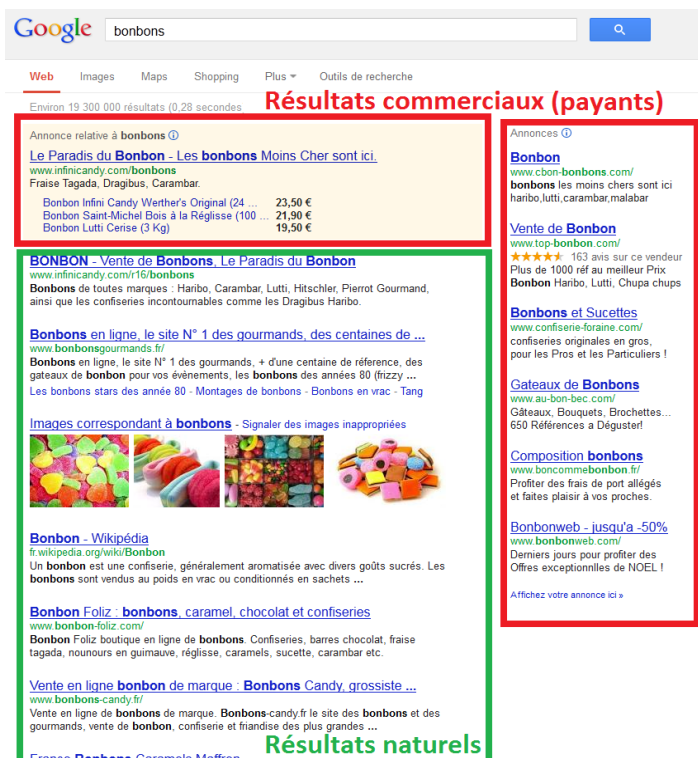
3. Comment faire pour améliorer le classement d'un site dans un moteur de recherche ?

Le référencement naturel utilise des techniques complexes basées sur la pertinence, la popularité, l'audience... des sites indexés. Difficile, voire impossible, d'avoir le contrôle sur le positionnement de son site web dans les réponses du moteur de recherche.

Pour améliorer le positionnement d'un site internet dans un moteur de recherche, **certaines entreprises achètent des emplacements publicitaires**. C'est ce que l'on appelle le **référencement payant**.

Vous remarquerez en effet que pour certaines requêtes, des résultats sponsorisés, (« annonces ») apparaissent sur le moteur de recherche.

Le référencement payant est une source de revenu importante, voire indispensable, pour les moteurs de recherche.



4. Est-ce que tout le contenu du web est référencé dans les moteurs de recherche ?

Vous ne le saviez peut-être pas mais il existe 3 sortes de web :

- Le World Wide Web : le web tel que vous utilisez de manière classique,
- Le Deep Web : le web profond,
- Le Dark Web : le web clandestin ou web sombre.

Seul le **World Wide Web** (www) est **indexé par les moteurs de recherche**. C'est ce que vous utilisez lorsque vous surfez tout naturellement. C'est également le domaine du Web qui fait l'objet d'une surveillance constante de la part des gouvernements du monde entier.

Le **Deep Web** (web profond) est l'ensemble des **sites internet** qui ne sont **pas indexés**, les **moteurs de recherche ne peuvent pas les trouver**. Le Deep Web contient tous nos dossiers médicaux, dossiers financiers, fichiers de médias sociaux et beaucoup d'autres informations importantes que nous voulons et que nous devons garder en sécurité.

Le **Dark Web** fait partie du Deep Web, mais il a une différence majeure : il a été **intentionnellement caché et est inaccessible aux navigateurs Web normaux**. Le contenu du Dark Web est divers, mais souvent illégal. C'est là que l'on trouve notamment les plateformes de ventes illégales, des groupes de piratage... mais tout n'est pas illégal sur le Dark Web, on y trouve aussi beaucoup de ressources qui y sont cachés pour lutter contre la censure et l'oppression de certains gouvernements.

