



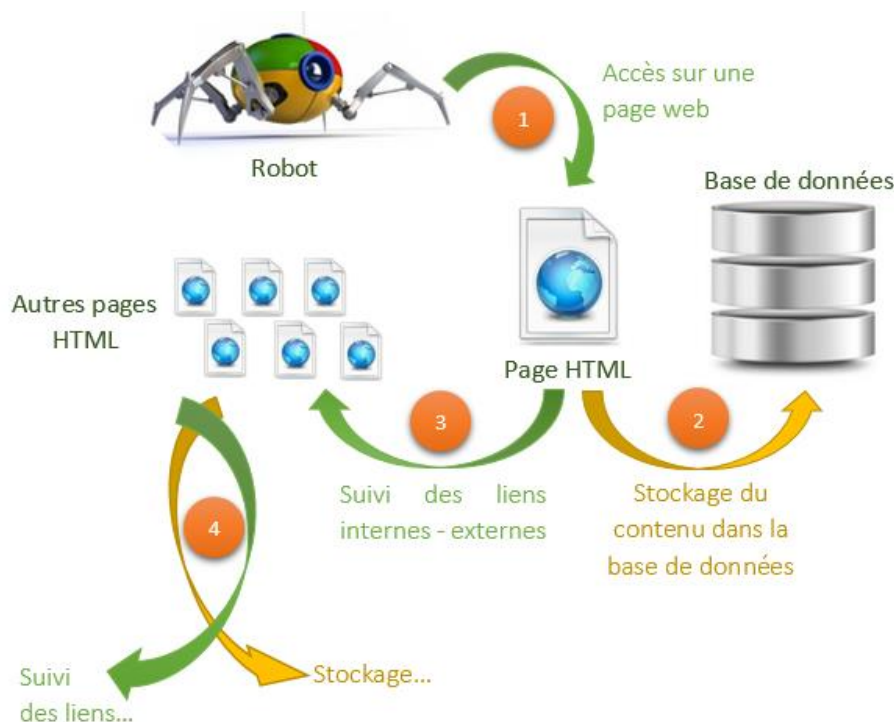
# Fonctionnement des moteurs de recherche

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, voir dont on ignore l'existence.**



## ➤ Comment fonctionnent-ils ?

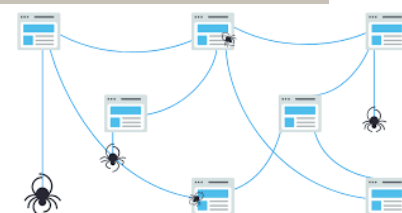
Des « robots » ou « web crawlers » ou « web spyders » (programmes informatiques) explorent automatiquement le Web. Ils sont conçus pour **collecter les ressources (pages Web, images, vidéos, documents Word, etc.) afin de permettre à un moteur de recherche de les indexer.**



Une fois les données collectées pas les robots, ces données vont être **indexées**, un **algorithme** va les classer en fonction de plusieurs critères, comme le nombre de liens pointant vers une page. Le principe de fonctionnement est fondé sur le fait que plus un site **est cité par d'autres sites, plus il sera considéré comme pertinent et donc plus son score sera élevé. Un bon score garantit une place de choix au site dans la page des résultats, c'est le référencement naturel**, appelé aussi SEO (Search Engine Optimization)



Il est à distinguer du **référencement payant** des annonces ou des liens sponsorisés. C'est ainsi que les moteurs de recherche gagnent de l'argent.





## ➤ Coté utilisateur...

### 1. Le « Deep Web »

Toutes les pages web, bien qu'accessibles avec un navigateur internet, **ne sont pas référencées par les moteurs de recherche**, car le développeur n'a pas codé la page pour qu'elle le soit. Ces pages ont des particularités : elles sont dynamiques, protégées par un mot de passe et contiennent des ressources volumineuses, entre autres. **Ces ressources non indexées par les moteurs de recherche composent le « Deep Web » (96% du Web).** Le Dark Web représente la partie illégale, seulement 6% du Deep Web. Les ressources indexées quant à elles, composent le Web de surface.



« Je ne suis pas un robot ! »

Les CAPTCHA servent à différencier les robots et les humains dans le cas où les pages se feraient espionner par un logiciel malveillant.



i'm not a robot

