

3 SOUS LE CAPOT DES MOTEURS DE RECHERCHE

Lorsqu'on lance une requête dans la barre du navigateur pour rechercher un article, une vidéo, etc., le navigateur interroge un moteur de recherche, qui doit nous proposer les résultats les plus pertinents possibles par rapport à notre demande. Comment fonctionne un moteur de recherche ? Ses réponses sont-elles toujours fiables ?

1 Parcourir l'immense graphe du Web

Le Web est un immense **graphe** : chaque page du Web est un nœud relié à d'autres nœuds par les liens hypertextes contenus dans la page. Les **moteurs de recherche** parcourent constamment ce graphe (c'est-à-dire toutes les pages Web) afin de collecter un maximum de contenus, qu'ils indexent. Grâce à cette indexation, un moteur de recherche pourra fournir rapidement à un internaute des réponses adaptées à sa requête. Il existe plusieurs moteurs de recherche, notamment Google (le plus utilisé), Bing de Microsoft, ou encore DuckDuckGo (qui ne collecte pas de données personnelles).

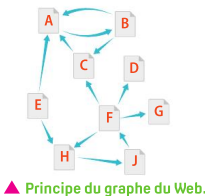
Principe du parcours du Web par le moteur de recherche

On utilise deux listes : les pages *déjà visitées* (aucune au départ) et les pages *à visiter* (par exemple, la page **E** au départ).

1. On prend une page parmi les pages à visiter, par exemple **E**.
2. On la transfère dans la liste des pages visitées, et on ajoute à la liste des pages à visiter les pages accessibles depuis **E** (ici, **A** et **H**) qui ne sont pas déjà visitées.
3. On continue jusqu'à ce qu'il n'y ait plus de pages à visiter.

VOCABULAIRE

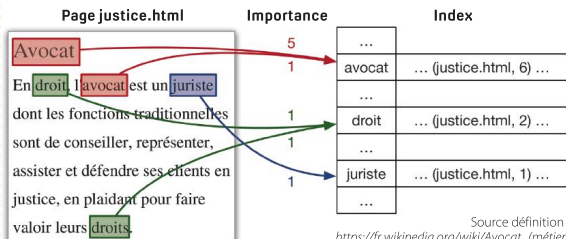
Graphe du Web
Graphe orienté dont les nœuds sont les pages Web, et les arcs les liens hypertextes entre les pages.
(► *Théorie des graphes*, p. 62)



2 Indexer pour pouvoir retrouver

Au fur et à mesure qu'il parcourt le Web, le moteur de recherche indexe les contenus : il crée un **index** qui associe à chaque mot l'ensemble des pages qui contiennent ce mot, avec l'importance du mot pour cette page. Exemples de critères d'importance d'un mot :

- un mot qui apparaît dans un titre est plus important (5 points) pour cette page que s'il apparaît dans un paragraphe (1 point) ;
- un mot qui apparaît souvent dans la page est plus important pour cette page que s'il n'apparaît qu'une fois (1 point par occurrence).



▲ Indexation de quelques mots de la page Web.

VOCABULAIRE

Moteur de recherche
Outil qui permet de trouver des ressources à partir d'une requête d'un utilisateur comportant des mots-clés.

Index
Dictionnaire associant à chaque mot les pages dans lesquelles ce mot apparaît.

Principe de l'indexation

Pour chaque mot **M** qui apparaît dans une page d'adresse **A** :

1. calculer son importance **I** ;
2. ajouter (**A**, **I**) à la liste des pages associées au mot **M**.

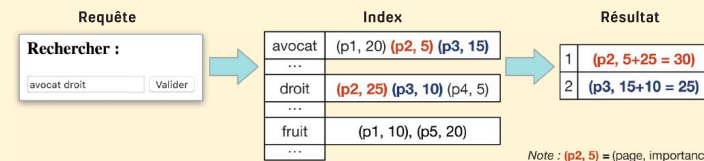
3 Répondre à une requête de recherche

Lorsque le moteur de recherche reçoit une requête, il utilise l'index pour trouver toutes les pages qui contiennent les mots-clés de la requête. Ensuite, il classe les pages par ordre de pertinence, en

combinant plusieurs facteurs :
– l'importance des mots (► *doc. 2*) ;
– la popularité (► *doc. 4*) ;
– le profil de l'utilisateur (► *p. 40*).



Ci-dessous, une recherche à partir des mots-clés « avocat » et « droit ». Afin de simplifier l'exemple, le classement des réponses est effectué seulement en fonction de l'importance des mots.



4 Les résultats sont-ils fiables ?

- La popularité d'une page dépend de plusieurs facteurs :
 - le nombre de pages qui ont des liens vers cette page (► *algorithme PageRank*, p. 51) ;
 - le nombre de fois que les utilisateurs choisissent cette page en réponse à une requête de recherche ;
 - l'achat de mots-clés au moteur de recherche par un site afin que ses pages soient mieux classées.
- Les moteurs de recherche combinent ces facteurs selon des formules qui sont gardées secrètes et qui changent régulièrement. C'est pourquoi les réponses à une même requête varient d'un moteur à l'autre.

Les premiers résultats de la requête « avocat » avec deux moteurs de recherche différents :

Moteur 1
Cabinet d'Avocats – Paris (pub)
Avocat (fruit) – Wikipédia
Avocat (métier) – Wikipédia

Moteur 2
Conseil Juridique (pub)
avocat / avocate - Onisep
Avocat (métier) - Wikipédia

ZOOM SUR...

L'optimisation d'un site

Elle consiste à choisir des mots-clés et à les placer de manière stratégique dans la page afin d'améliorer son classement. Par exemple, pour la page d'un groupe de musique pop, on pourrait placer les mots « musique », « pop », « groupe » dans les titres. Certains sites tentent de tromper les moteurs de recherche en utilisant des textes cachés ou en achetant des liens pour augmenter la popularité de leurs pages.

QUESTIONS

- 1 **Doc. 1** Continuez le parcours du graphe à partir du nœud **E** en listant les pages visitées et les pages à visiter à chaque étape. Combien d'étapes sont nécessaires pour visiter tout le graphe ?
- 2 **Doc. 1** Que se passe-t-il si l'on commence le parcours par le nœud **A** ?
- 3 **Doc. 2** Complétez l'index avec les mots « avocat » et « fruit » des pages HTML

suivantes en appliquant les critères d'importance :

- avocat.html **Avocat**
L'avocat est le fruit de l'avocatier, un arbre originaire du Mexique.
- fruit.html **Fruit**
Un fruit est un aliment végétal, à la saveur sucrée, généralement consommé cru.
- 4 **Doc. 3** Calculez le classement des résultats pour la requête « avocat droit » en supposant que l'importance

- du premier mot-clé compte double. Qu'observez-vous ?
- 5 **Doc. 3** Calculez et comparez les résultats des requêtes de recherche avec les mots-clés : « avocat fruit » ; « avocat droit » ; « avocat ».
 - 6 **Doc. 4** Effectuez la même requête sur différents moteurs de recherche (Google, Bing, DuckDuckGo, etc.) et comparez les résultats.