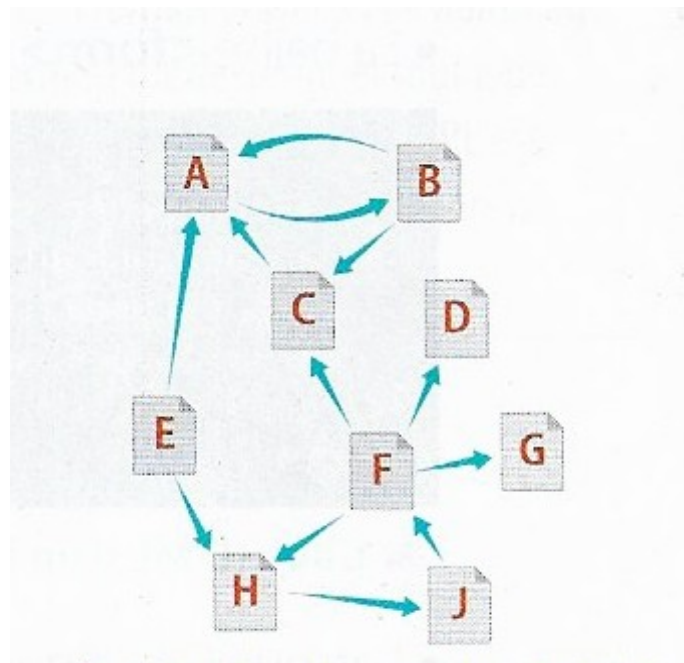


## Les moteurs de recherche.

Lorsqu'on lance une requête dans la barre d'un navigateur pour rechercher un article, une vidéo, ..., 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 ?

Document 1 : Le Web est un immense graphe : chaque page Web est un nœud relié 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, Bing ou DuckDuckGo.

Principe du parcours su 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 accessibles depuis E (ici A et H) qui ne sont pas encore visitées.
3. On continue jusqu'à ce qu'il n'y ait plus de pages à visiter.

Question 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 ?

Question 2 : Que se passe-t-il si l'on commence le parcours par le nœud A ?

Document 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 pour d'un mot :

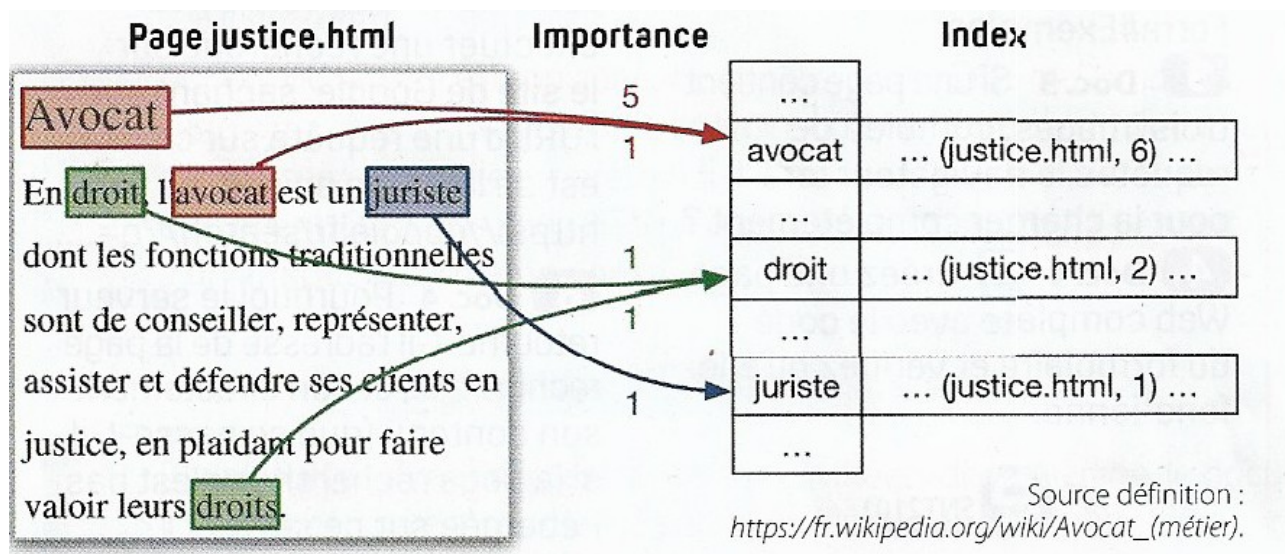
- un mot qui apparaît dans un titre est plus important (5 points) pour cette page que s'il apparaît dans un paragraphe (1 point)
- un mot qui apparaît souvent dans la page est plus important pour cette page que s'il n'apparaît qu'une fois (1 point par occurrence).

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 liste des pages associées au mot M.

Exemple :



Question 3 :

Voici deux exemples de page html :

page avocat.html

### Avocat

L'avocat est le fruit de l'avocatier, un arbre originaire du Mexique.  
L'avocat peut être utilisé pour faire du guacamole.

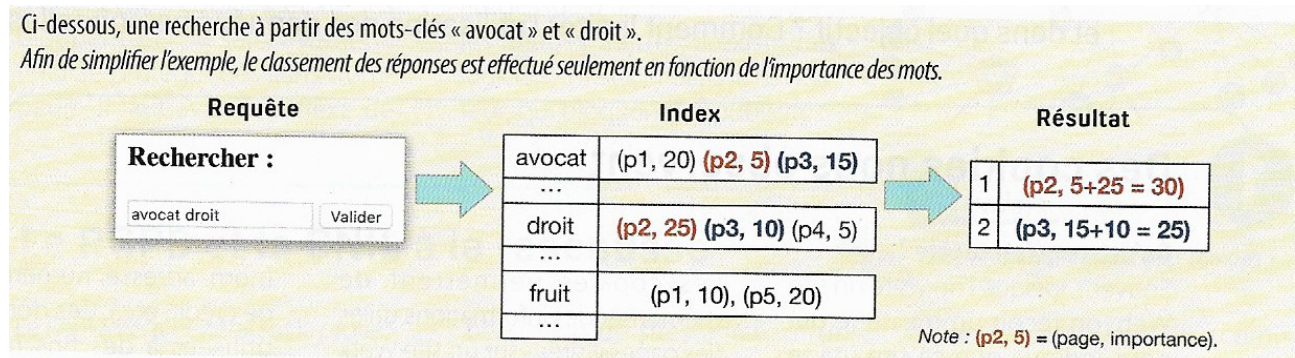
## Fruit

Un fruit est un aliment végétal, à la saveur sucrée, généralement consommé cru.

Compléter l'index avec les mots « avocat » et « fruit » des deux pages ci-dessus.

Document 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 tels que l'importance des mots, la popularité, le profil de l'utilisateur.



Question 4 : Calculer le classement des résultats pour la requête « avocat droit » en supposant que l'importance du premier mot-clé compte double. Qu'observez vous ?

Question 5 : Calculer et comparer les résultats des requêtes de recherche avec les mots-clés : « avocat droit », « avocat fruit », « avocat ».

Document 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.
- Le nombre de fois que les utilisateurs choisissent cette page en réponse à une requête de recherche.
- L'achat des 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 algorithmes qui sont gardées secrets et qui changent régulièrement. C'est pourquoi les réponses à une même requête varient d'un moteur à l'autre.

Question 6 : Effectuez la même requête sur différents moteurs de recherche et comparez les résultats. Quelles explications pouvez vous fournir aux différences trouvées ? Question 7 : Quels conseils pourriez vous donner à une personne désirant mettre son site en ligne afin qu'il soit le plus visible possible?