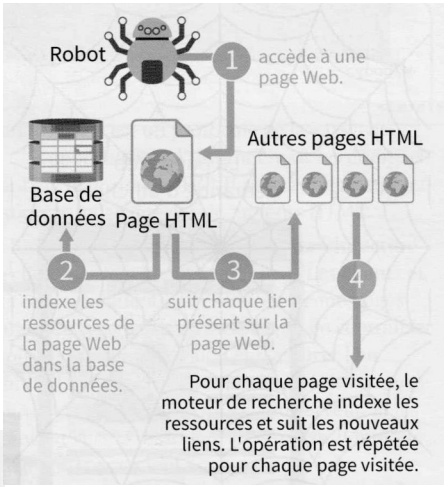
Les moteurs de recherche

1- L'indexation

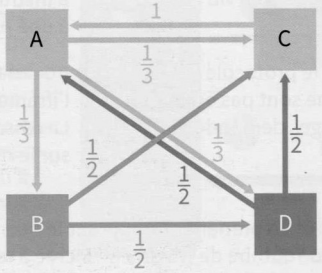
Des « robots » ou « crawler » (programmes informatiques) parcourent le Web et indexent les pages à partir des mots qu'elles contiennent et en lien avec leur adresse URL. Par exemple, l'index peut indiquer que le mot « Louvre » est utilisé sur les pages 10, 27 et 157. Cela permet de gagner du temps de réponse face à la requête du visiteur. Toutes les pages ne sont pas sauvegardées. Certaines pages provenant de sites illégaux sont tout simplement blacklistées.



2- Le classement

Pour calculer un score, l'algorithme :

1. part du principe que chaque site possède un « vote » égal à 1. Ce vote se répartit également sur le nombre de liens sortants (qui pointent vers d'autres sites) ;
2. additionne le score total de chaque site ;
3. fait un classement.



Le score du site A est de 2,5, celui du site B de 1,33, celui du site C de 2,33 et celui du site D de 1,83. On obtient donc le classement ou l'ordre d'affichage suivant :
site A – site C – site D – site B.

Une fois les données collectées par les robots, 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**.

3- Tout est référencé ?

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 **Web profond** ou le **« Deep Web »** (96 % du Web). Le Dark Web en représente la partie illégale. Les ressources indexées quant à elles, composent le **Web de surface**.

Questions :

1) Les moteurs de recherche recensent-ils toutes les pages du web ?

Dans l'exemple 2), le site C ajoute un lien vers le site D et le site B retire le sien vers le D.

2) Refaire le graphique pondéré puis calculer le qcore de chaque site

3) Proposer les noms de 4 autres moteurs de recherche que google.fr et bing.fr

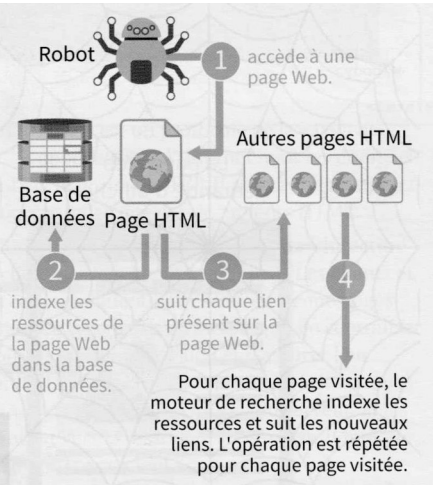
4) Pensez vous que chaque moteur réalise lui-même sa propre indexation ?

5) Proposer le nom de 4 navigateurs internet que l'on peut utiliser pour naviguer sur le web

Les moteurs de recherche

1- L'indexation

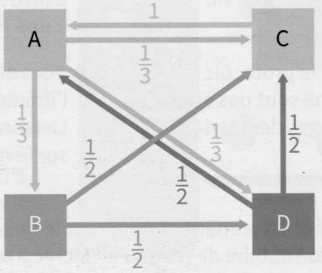
Des « robots » ou « crawler » (programmes informatiques) parcourent le Web et indexent les pages à partir des mots qu'elles contiennent et en lien avec leur adresse URL. Par exemple, l'index peut indiquer que le mot « Louvre » est utilisé sur les pages 10, 27 et 157. Cela permet de gagner du temps de réponse face à la requête du visiteur. Toutes les pages ne sont pas sauvegardées. Certaines pages provenant de sites illégaux sont tout simplement blacklistées.



2- Le classement

Pour calculer un score, l'algorithme :

1. part du principe que chaque site possède un « vote » égal à 1. Ce vote se répartit également sur le nombre de liens sortants (qui pointent vers d'autres sites) ;
2. additionne le score total de chaque site ;
3. fait un classement.



Le score du site A est de 2,5, celui du site B de 1,33, celui du site C de 2,33 et celui du site D de 1,83. On obtient donc le classement ou l'ordre d'affichage suivant :
site A – site C – site D – site B.

Une fois les données collectées par les robots, 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**.

3- Tout est référencé ?

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 **Web profond** ou le **« Deep Web »** (96 % du Web). Le Dark Web en représente la partie illégale. Les ressources indexées quant à elles, composent le **Web de surface**.

Questions :

1) Les moteurs de recherche recensent-ils toutes les pages du web ?

Dans l'exemple 2), le site C ajoute un lien vers le site D et le site B retire le sien vers le D.

2) Refaire le graphique pondéré puis calculer le qcore de chaque site

3) Proposer les noms de 4 autres moteurs de recherche que google.fr et bing.fr

4) Pensez vous que chaque moteur réalise lui-même sa propre indexation ?

5) Proposer le nom de 4 navigateurs internet que l'on peut utiliser pour naviguer sur le web