

TP 1 : Les moteurs de recherche**1. Navigateur ou moteur ?**

Un navigateur web est un logiciel familier mais d'une extrême complexité, comportant des millions de lignes de code. Il en existe plusieurs.

Question 1 : Saurez-vous identifier les navigateurs suivants en éliminant les intrus ?

il en existe bien d'autres ...

Au besoin, utiliser internet pour répondre aux questions suivantes :

Question 2 : Dans le tableau, identifier les navigateurs web. Voir tableau ci-dessus

Question 3 : Dans le tableau, identifier les moteurs de recherche.

Question 4 : Quel logiciel utiliser pour naviguer sur le web avec un smartphone ?

Question 5 : (Synthèse) : Quelles différences entre un navigateur et un moteur de recherche, un annuaire web?

2. Protection vie privée

Question 6 : Qu'est-ce qu'un cookie ?

Question 7 : Dans la barre de recherche quelle est la différence entre http et HTTPS ?

Question 8 : Qu'est ce que le mode navigation privée ?

3. Référencement ou « page ranking »

3.1 Exercice 1 :

En choisissant **3 à 4 moteurs de recherche différents**, effectuez une même recherche (même mots clés !).

1. Mettez vous en groupe de 2 élèves et observez les 3 premières réponses d'une recherche effectuée avec les mêmes mots clés.
2. Présentez le tout sur une page A4 en utilisant les différents outils :
 - Outil capture
 - Paint ...

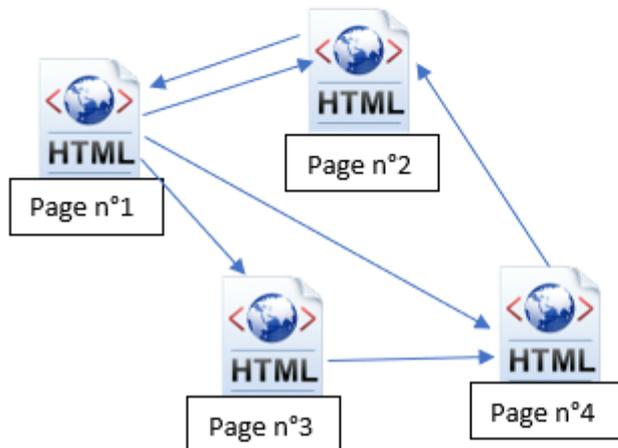
On se rend compte que le résultat d'une même recherche varie d'un moteur à l'autre, et cela est du au « page ranking »

- Le **PageRank** est l'algorithme d'analyse des liens hypertextes utilisé pour le classement des pages Web par le moteur de recherche Google.
- Le **PageRank** n'est qu'un indicateur parmi d'autres dans l'algorithme qui permet de classer les pages du Web dans les résultats de recherche. Ce système a été inventé par **Larry Page, cofondateur de Google. Ce mot est une marque déposée.**
- Le principe de base est d'attribuer à chaque page une valeur proportionnelle au nombre de fois que passerait par cette page un utilisateur parcourant le Web en cliquant aléatoirement, sur un des liens apparaissant sur chaque page.
- A chaque instant, des "robots", appelés "spiders", parcourent la toile, de lien en lien, et mettent ainsi sans cesse à jour le PageRank des pages du web.

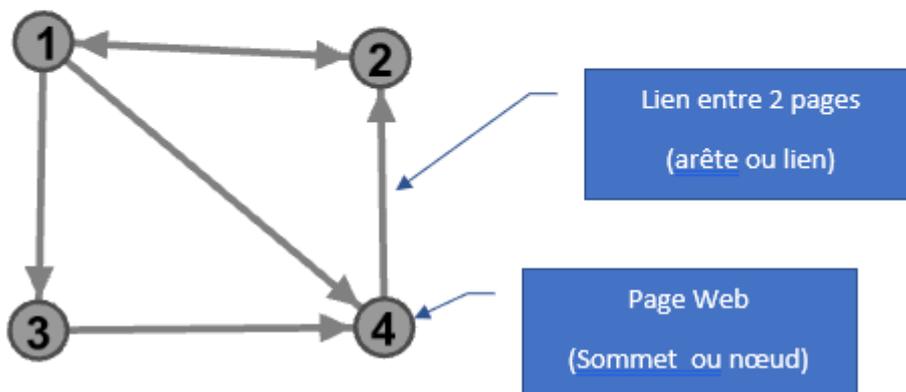
3.2. Principe

Représentation des liens des pages Web entre elles par un graphe

Tu as vu qu'un site Web pouvait contenir des liens hypertextes. C'est-à-dire des connexions vers d'autres pages Web du même site ou vers d'autres sites Web. Ces liens sont donc orientés. Un lien de la page Web 1 vers page Web 2, n'est pas du tout la même chose qu'un lien de la page Web 2 vers page Web 1 ! Il faut donc mettre une flèche sur ce lien.



Ces liens entre pages Web peuvent être représentés par un graphe qui relie des sommets (nœuds) entre eux par des arêtes (liens).



→ Pour classer les résultats de recherche on peut imaginer différentes possibilités, par exemple faire voter les internautes ou recourir à un expert du domaine. Mais ces solutions faisant appel à un être humain ont plusieurs défauts : comment être sûr de l'impartialité des personnes ? Comment faire lorsqu'une nouvelle page apparaît ? Sans compter le temps que cela nécessite. Bref... c'est pas top.

→ C'est en partant de l'idée d'améliorer le classement des recherches que Larry Page et Sergey Brin, fondateurs de Google, décident de créer le « PageRank ». Il s'agit d'un nombre, compris entre 0 et 1, déterminé sans intervention d'un être humain et qui classe les pages de la « meilleure » à « la moins bonne ». Ce PageRank obéit à deux règles :

Règle 1 : Le PageRank attribué à une page doit être d'autant plus élevé que celle-ci est référencée dans une page faisant autorité (dont le score élevé).

Règle 2 : Le PageRank attribué à une page doit être d'autant moins élevé que celle-ci est référencée dans une page contenant un grand nombre de références.

La règle 1 considère donc que si un site bien classé fait un lien vers un autre site, c'est que cet autre site doit également être bon. La règle 2 permet d'éliminer les sites qui ne feraient que faire des liens et évite qu'un site contenant des millions de liens mais aucune information ne soit bien classé.

Reste un problème : les idées c'est bien, mais comment les mets-on en place ? Larry Page et Sergey Brin sont tous deux étudiants en mathématiques et informatique et ils ont compris que les cours théoriques d'algorithmique qu'ils ont suivis allaient leur permettre de révolutionner la recherche sur le web. Ils décident d'appliquer un procédé qui peut sembler à priori peu efficace : le hasard !

3.3. Le surfeur aléatoire

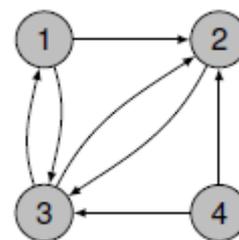
Les deux hommes réalisent un programme, qu'on nomme maintenant le surfeur aléatoire, qui va accomplir les tâches suivantes :

- **Etape 1** : Quand un internaute fait une recherche sur un mot-clé il fait la liste de tous les sites relatifs à ce mot-clé (mais il ne classe rien à cette étape)
- **Etape 2** : Il fait ensuite la liste de tous les liens entre les sites de sa liste
- **Etape 3** : Il choisit un site au hasard et parcourt les liens en comptant combien de fois il visite chaque site
- **Etape 4** : A la fin de son parcours il classe les sites de celui qu'il a le plus visité à celui qu'il a le moins visité

Bon ben... y a plus qu'à vérifier si cette idée bizarre peut fonctionner. Et là c'est à vous de jouer : commencez par mettre vous mettre en îlots de quatre, lisez ensuite les règles du jeu et une fois que vous êtes sûrs d'avoir compris, appelez le professeur pour lui expliquer les règles. Si vous avez bien compris il vous donnera le dé pour commencer.

Règles du jeu : on imagine qu'un internaute a fait une recherche sur un mot-clé. En réalisant l'étape 1 le surfeur a trouvé quatre sites relatifs à ce mot clé (on les note 1, 2, 3 et 4). En réalisant l'étape 2 il a trouvé que :

- Le site 1 a des liens vers les sites 2 et 3
- Le site 2 a un lien vers le site 3
- Le site 3 a des liens vers les sites 1 et 2
- Le site 4 a des liens vers les sites 2 et 3



Pour visualiser cela plus simplement on a fait un graph ci-contre :

A l'étape 3 vous devez choisir un point de départ au hasard, pour cela lancez le dé. Si vous obtenez 1 vous commencerez sur le site 1, si vous obtenez 2 vous commencerez sur le site 2 etc. Si vous obtenez 5 ou 6, recommencez jusqu'à obtenir 1 ou 2 ou 3 ou 4.

Placez votre pion sur le site de départ et ajoutez une visite au site concerné dans le tableau ci-dessous :

Site n°	1	2	3	4
Nombre de visite				

Lancez ensuite le dé :

- Si vous obtenez 5 ou 6, recommencez sans compter l'essai
- Si vous obtenez 1, 2, 3 ou 4 : regardez le graph. Pouvez-vous vous déplacer en suivant un lien ?
 - Si oui, déplacez le pion et ajoutez une visite au site dans le tableau
 - Si non, restez sur le site actuel et relancez le dé
- Recommencez 30 fois

Classez les sites du meilleur au moins bon selon le surfeur : _____

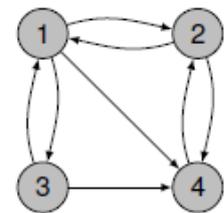
Notez les résultats des autres groupes de la classe : _____ ; _____ ; _____ ; _____

_____ ; _____ ; _____ ; _____ ; _____ ; _____ ; _____ ; _____

Alors ça donne quoi ce surfeur aléatoire ? _____

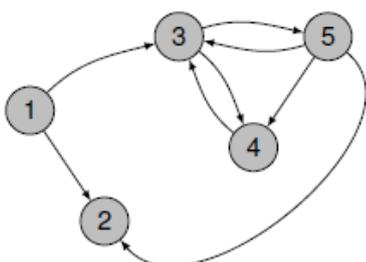
On vérifie que ce n'est pas un coup de chance ?

Site n°	1	2	3	4
Nombre de visite				

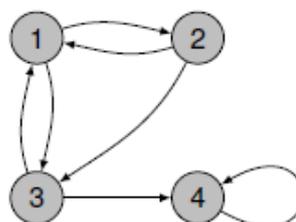


Conclure : _____

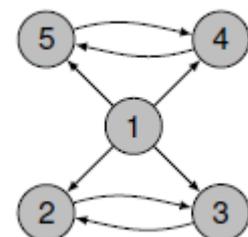
4. Les limites du surfeur



Graph A



Graph B



Graph C

Les trois graphes ci-dessus représentent chacun un problème qui bloque le surfeur. Pour chaque graph, et sans faire de lancer de dé, trouvez-le et expliquez-le : _____

Proposer des solutions qui permettraient de résoudre ces trois problèmes : _____

5. La solution de Larry Page et Sergey Brin : les mathématiques encore et toujours

Etudiants studieux les deux fondateurs de Google connaissent le théorème Peyron-Froebenius qui précise que si de chaque sommet part un lien vers chacun des autres sommets alors les fréquences des positions au cours de notre surf aléatoire vont toujours converger vers la même valeur et cela indépendamment de la position de départ. Ainsi si une page ne comporte aucun lien vers l'extérieur, comme le sommet 2 du graphe A, on crée artificiellement un lien de la page vers toutes les autres pages.

L'idée de génie : Larry Page et Sergey Brin modifie le code du surfeur pour qu'à chaque étape il y a 85% de chances de procéder comme on l'a fait précédemment et 15% de chances de faire un saut aléatoire vers n'importe quel sommet (ils ont démontré auparavant que cela rentrait dans le cadre du théorème Peyron-Froebenius).

Vous disposez maintenant d'un dé à 6 faces et d'un dé à 100 faces. Comment faire pour mettre en œuvre l'idée précédente ? _____

Grace à cette méthode proposer un classement des pages des quatrième et cinquième graphes :

Cette méthode modifie-t-elle le classement de pages pour lesquelles il n'y avait pas de problèmes ?
