

**Activité sur les moteurs de recherche**

1. Citer les moteurs de recherche que vous connaissez
  
2. Répondez aux questions suivantes après avoir regardé cette vidéo <https://youtu.be/OA5fQER40Wg> :
  - a. Expliquer la nécessité du développement des moteurs de recherche dans les années 1990.
  
  - b. Qui sont les fondateurs de Google ?
  
  - c. Expliquer le fonctionnement du moteur de recherche Google
  
  - d. Rechercher sur un moteur de recherche quel est l'autre nom donné au programme qui parcourt les pages web pour un moteur de recherche ?

3. Le PageRank

En 1998, deux jeunes doctorants de l'université de Stanford publiaient un article intitulé "The PageRank Citation Ranking: Bringing Order to the Web" présentant les résultats d'un nouvel algorithme permettant de classer les pages web selon leur popularité et montrant la précision de cet algorithme sur un nouveau moteur de recherche appelé ... Google !

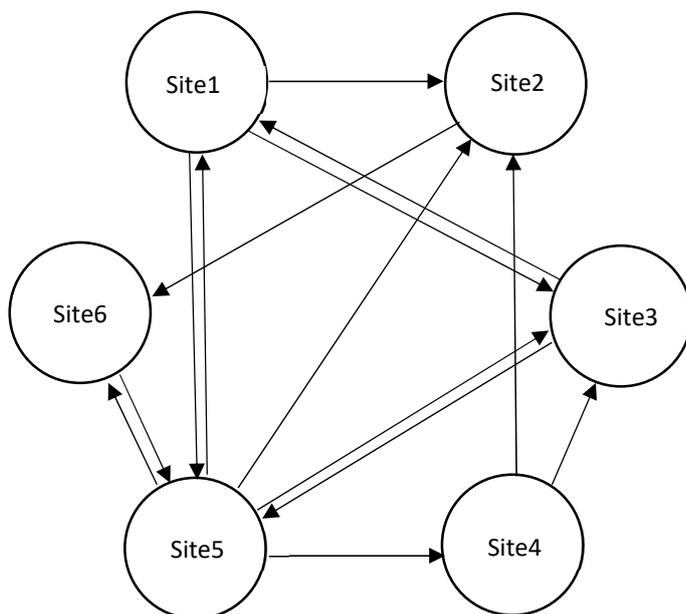
- On simule le comportement d'un internaute.
- On démarre au hasard sur une page
- On suit de façon aléatoire un lien se trouvant sur cette page
- Et ainsi de suite

On impose le fait que dans 15% des cas , l'utilisateur abandonne sa navigation pour repartir d'une page au hasard.

Notre utilisateur va se balader de pages en pages, en faisant son petit tour du net. A chaque fois que l'utilisateur tombe sur une page donnée, cette page gagne un point. Et à la fin, la page ayant le plus de point est alors la page la plus populaire du réseau !

Considérons 6 sites numérotés 1, 2, 3, 4, 5, 6 ayant des liens hypertexte entre eux.

- Le site1 contient un lien vers les sites "site2", "site3" et "site5" (3 liens sortants)
- Le site2 contient un lien vers .....
- Le site3 contient un lien vers .....
- Le site4 contient un lien vers .....
- Le site5 contient un lien vers .....
- Le site6 contient un lien vers .....



D'après le graphique quel doit être le site qui a le PageRank le plus élevé et qui peut sortir en tête des résultats de recherche ? Justifier

Simulation d'un PageRank : ouvrir Edupython (ou Spyder, ou autre) et copier le code suivant :

```
import random
# création d'un dictionnaire de liens Hypertextes
Hypertext = {}
# création d'un dictionnaire pour le nombre de visites par site
Walk_Number = {}
# une variable pour le nombre total de visite
Total_Walk = 0
#liste des sites web
Websites = ["site1","site2","site3","site4","site5","site6"]
# les liens hypertextes
# le dictionnaire possède des clés ( nom des sites)
# Qui contiennent des listes (liens hypertextes)
Hypertext["site1"] = ["site2","site3","site5"]
Hypertext["site2"] = ["site6"]
Hypertext["site3"] = ["site1","site5"]
Hypertext["site4"] = ["site2","site3"]
Hypertext["site5"] = ["site1","site2","site3","site4","site6"]
Hypertext["site6"] = ["site5"]
print(Hypertext)
# On initialise à 0.0 les visites des sites
Walk_Number["site1"] = 0.0
Walk_Number["site2"] = 0.0
Walk_Number["site3"] = 0.0
Walk_Number["site4"] = 0.0
Walk_Number["site5"] = 0.0
Walk_Number["site6"] = 0.0

i = 0
while i < 1000:
    x = random.choice(Websites)
    while random.random() < 0.85:
        Walk_Number[x] = Walk_Number[x] + 1
        Total_Walk = Total_Walk + 1
        x = random.choice(Hypertext[x])
    i = i + 1
print (Walk_Number)
print (Total_Walk)
```

Questions :

- Que représentent les lignes 13 à 18 ?
- Sachant que `random.random()` renvoie un nombre au hasard entre 0 et 1, à quel point de la simulation correspond la ligne 31 ?
- Que contient à la fin de l'exécution du programme le dictionnaire (sorte de tableau) `Walk_Number` ?
- Quel est le site ayant le PageRank le plus élevé ? Cela correspond-il au résultat attendu ?
- Le webmestre du site2 ajoute un lien vers le site3. Comment est modifié le PageRank du site3 ? Quelles modifications faudrait-il faire dans le programme pour tenir compte de cette modification ? Quel est le PageRank obtenu pour le site3 ?

#### 4. Les mots-clés

Remarque : taper à chaque fois l'expression comprise entre « et »

Afin de préparer un exposé sur l'escargot de Bourgogne, une élève fait une recherche dans Google en tapant dans la barre de recherche : «escargot de Bourgogne». Elle ne souhaite pas consulter de recettes d'escargots.

Les résultats obtenus sont-ils différents pour une recherche correspondant à : escargot bourgogne ou encore à : escargot Bourgogne ? Que peut-on en conclure ?

Suivre le lien d'un résultat affiché par Google. Appuyer sur F12 pour démarrer les outils de développement.

A quelle partie du site correspond la première ligne affichée par google ? (Rechercher dans la partie <head> du code...)

Le descriptif du site affiché par google apparaît sur le site après un texte correspondant aux mots clés de la recherche effectuée. Quelles balises entourent ce texte sur le site ? On pourra s'aider de cette icône  permettant de sélectionner un texte du site et de colorier la partie correspondante du code.

Pour affiner sa recherche l'élève tape dans la barre de recherche «"escargot de bourgogne"». Que peut-on dire des résultats qu'elle obtient ?

Pour se passer des recettes elle essaie «escargot bourgogne -recette». Que peut-on dire des résultats obtenus ?

Elle essaie : «+escargot +bourgogne +helix». Que peut-on dire des résultats obtenus ?

D'après la publicité de Qwant visible à cette adresse : <https://youtu.be/Evahh1PXJlg> quel est le principal problème de Google ?

5. Faire une recherche pour estimer le coût énergétique d'une recherche faite sur le web. Argumenter votre réponse en fournissant les adresses des sites consultés.