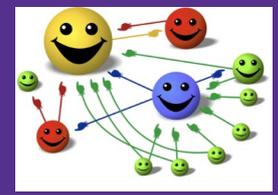


COMMENT FONCTIONNE UN MOTEUR DE RECHERCHE 2 ?



Indexation, pagerank,

Nom :

Prénom :

Classe :

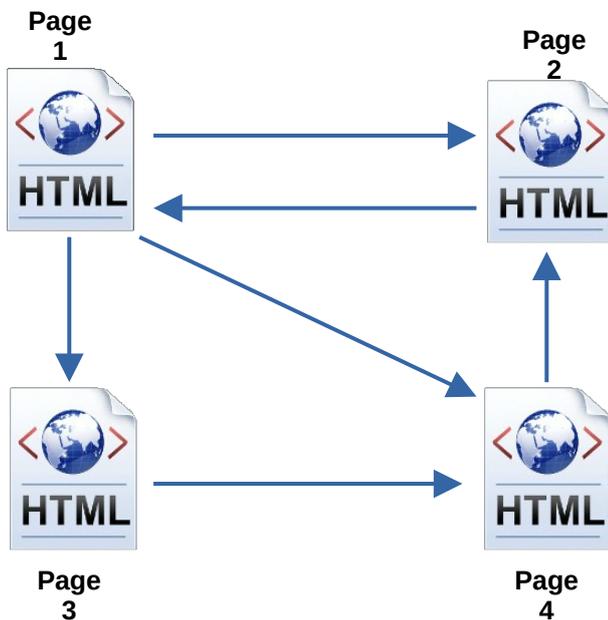
Date :

Situation

Tu as pu voir que pour qu'un moteur de recherche puisse classer des pages Web et proposer une page Web pertinente par rapport à la requête, le calcul de l'occurrence d'un ou des mots clé n'était pas suffisant.

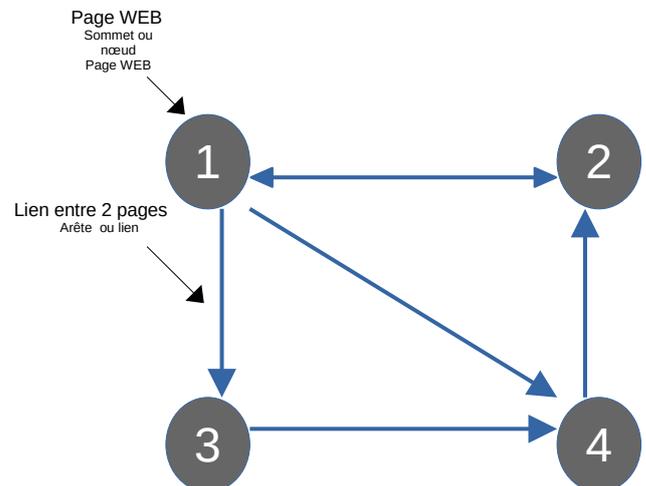
Tu as pu découvrir qu'une fois l'indexation effectuée, un nouvel algorithme améliore le classement des pages WEB.

Graphe de représentation des liens entre pages



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).



Le jeu de simulation

On peut logiquement penser que plus une page Web est référencée par d'autres pages Web, plus cette page est importante, populaire, et donc bien classée dans un moteur de recherche.

Autrement dit, un lien d'une page A vers une page B peut être considéré comme un vote en disant que la page A vote pour la page B. Mais toutes les pages n'ont pas le même pouvoir de vote. Pour finir, chaque page peut avoir plus ou moins d'importance suivant un nombre compris entre 0 (pas du tout importante) et 1 (très importante).

On peut trouver cette importance en imaginant un surfeur qui navigue sur un ensemble de pages Web aléatoirement. Il compte alors le nombre de fois qu'il passe par les mêmes pages Web.

On imagine bien que la page Web la plus visitée aléatoirement semble la plus « référencée ».

Utilise un dé ou l'application en ligne « <https://www.de-en-ligne.fr/1-des/4-faces.html> »_Et choisi un dé octogonal avec 3 chiffres et un dé à 2 chiffres.



Règle du jeu.

Lance le dé, et déplace-toi dans le graphe vers la page tirée (attention de bien respecter les liens orientés du graphe). Si cette page est accessible, alors tu positionnes un 1 dans la case de la page. Tu effectueras 40 lancés de ton dé. **Il sera parfois peut-être inutile de lancer le dé, car il n'y a qu'un seul lien entre deux pages Web !**

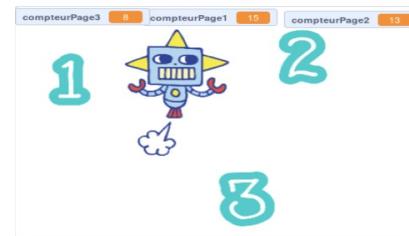
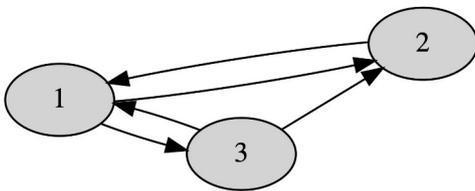
Complète le tableau suivant :

Avec les dés

Page Web visitée	Page Web 1	Page Web 2	Page Web 3	Page Web 4
Nombre de fois où je tombe sur cette page web au hasard				
Fréquence				

En comparant ces valeurs à celles trouvées avec les autres groupes , que constates-tu ?

Voici un programme d'une simulation incomplète en Scratch dans le fichier PageRankEleve.sb3 et sa traduction en Python dans le fichier PageRankEleve.py.



```

quand cliqué
mettre page à 1
mettre compteurPage1 à 0
mettre compteurPage2 à 0
mettre compteurPage3 à 0
répéter 1000 fois
attendre 1 secondes
si page = 1 alors
    si nombre aléatoire entre 1 et 2 = 1 alors
        aller à Glow-2
        mettre page à 2
        mettre compteurPage2 à compteurPage2 + 1
    sinon
        // Le robot est en page 1 et va vers la page 3
sinon
    si page = 2 alors
        aller à Glow-1
        mettre page à 1
        mettre compteurPage1 à compteurPage1 + 1
    sinon
        // Le robot est en page 3
    
```

```

#Ceci est un commentaire
#randint(a,b) retourne un entier aléatoire entre a et b
from random import randint

page = 1
compteurPage1 = 0
compteurPage2 = 0
compteurPage3 = 0
for k in range(1000):
    if page == 1:
        if randint(1, 2) == 1:
            page = 2
            compteurPage2 = compteurPage2 + 1
        else:
            "Le robot est en 1 et va vers 3"
    else:
        if page == 2:
            page = 1
            compteurPage1 = compteurPage1 + 1
        else:
            "le robot est en page 3"
print(compteurPage1, compteurPage2)
    
```

Sur quelle page se trouve le robot au début du programme ?

Ouvrir le fichier PageRankEleve.sb3 avec Scratch et le fichier PageRankEleve.py avec Pyzo, environnement de programmation Python.

Comment est traduite, en Python, l'instruction Scratch ?



Comment s'appelle cette structure algorithmique ?

On donne ci-dessous un extrait de code Scratch et sa traduction en Python :



```
if page == 1:
    if randint(1, 2) == 1:
        page = 2
        compteurPage2 = compteurPage2 + 1
    else:
        "Le robot est en 1 et va vers 2"
```

Expliquer le code Scratch. Pourquoi parle-t-on de déplacement aléatoire du robot ?

Quelle est la correspondance en Scratch de l'instruction `compteurPage2 = compteurPage2 + 1` ? Que représente `compteurPage2` ?

Comment les blocs d'instructions sont-ils délimités en Python ? Comment cela s'appelle-t-il ?

Compléter le bloc Scratch du Sinon correspondant au cas où le robot est en page 2. Traduire en parallèle dans le programme Python.



```
else:
    if page == 2:
        page = 1
        compteurPage1 = compteurPage1 + 1
    else:
        "le robot est en page 3"
```

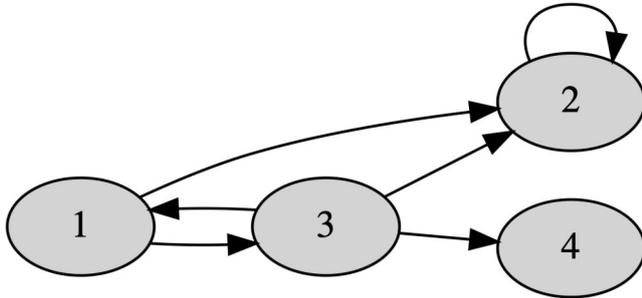
Exécuter les programmes Scratch et Python lorsqu'ils sont complets. Le nombre de passages du robot sur chaque page représente son PageRank. Quel classement de notoriété des trois pages obtenez-vous après simulation de mille déplacements aléatoires du robot d'indexation ? Comparez avec votre voisin. Ce résultat était-il prévisible ?

Avec les simulations logiciels				
Page Web visitée	Page Web 1	Page Web 2	Page Web 3	Page Web 4
Page Rank (mblock)				
Page Rank (python)				

Quelle limite vois-tu à ce classement ? Passe à la question suivante pour te donner une idée

Quel problème risque de se poser si on veut simuler le PageRank avec le graphe de documents suivant qui comporte 4 pages reliées par des liens hypertextes ?

Quelle solution pourrait-on imaginer pour contourner cet obstacle aux déplacements du robot de page en page ?



Ouvrir le fichier Liens-Page Rank avec le simulateur de graphe Gephi. Il permet de calculer le PageRank de chaque page. Remplir le tableau.

Avec la formule				
Page Web visitée	Page Web 1	Page Web 2	Page Web 3	Page Web 4
Page Rank				

Pour aller plus loin ...

A partir de la vidéo répondre aux questions :

Qu'est-ce que le facteur d'atténuation :

Quel facteur prenons nous pour calculer le Page Rank ?

En utilisant la formule :

$$PR(P) = \frac{0.85}{1-d} + d \sum_{P_i \rightarrow P} \frac{PR(P_i)}{L(P_i)}$$

Calculer le Page Rank de chacune des pages Web

Avec la formule				
Page Web visitée	Page Web 1	Page Web 2	Page Web 3	Page Web 4
Page Rank				

Quelle est la signification de l'acronyme SEO ?