

TD - Thème: Le web



voir page 50 de votre livre

Objectif

Simuler avec python le classement des pages sur internet (pagerank)

Table des matières

[1 Première étape: simuler aléatoirement l'accès à 6 pages web](#)

[1.1 Premier principe](#)

[1.2 Mise en place](#)

[1.3 Comprendre le script](#)

[2 Seconde étape: sites avec liens \(backlinks\)](#)

[2.1 Second principe](#)

[2.2 Créer la table des redirections par hyperliens](#)

[2.3 Créer le second script](#)

[2.4 Répondre aux questions](#)

[2.5 Votre script](#)

[3 Troisième étape: L'influenceur](#)

[3.1 Troisième principe](#)

[3.2 Mes simulations](#)

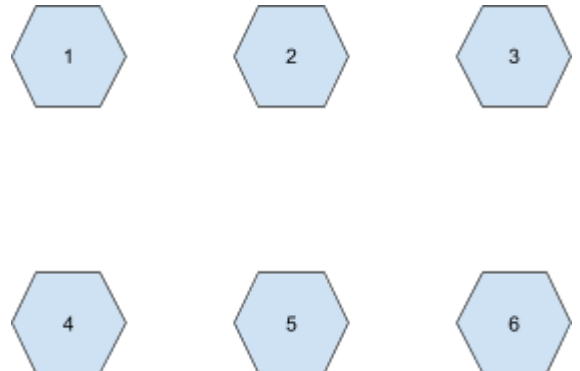
[3.3 Votre script](#)

[3.4 Répondre aux questions](#)

1 Première étape: simuler aléatoirement l'accès à 6 pages web

1.1 Premier principe

Simuler 1000 internautes qui atteignent aléatoirement 6 sites différents. Calculer la fréquence pour chaque site.



1.2 Mise en place

Copier /coller ce script

```
# début du script
from random import randint

site1,site2,site3,site4,site5,site6=0,0,0,0,0,0
n=0

for i in range (1000):
    tirage = randint(1,6)
    n=n+1
    if tirage==1:
        site1=site1+1
    elif tirage==2:
        site2=site2+1
    elif tirage==3:
        site3=site3+1
    elif tirage==4:
        site4=site4+1
    elif tirage==5:
        site5=site5+1
    else:
        site6=site6+1

print(site1,site2,site3,site4,site5,site6)
print("Total= ",site1+site2+site3+site4+site5+site6)
print("pagerank site1= ",site1/n)
print("pagerank site2= ",site2/n)
print("pagerank site3= ",site3/n)
print("pagerank site4= ",site4/n)
print("pagerank site5= ",site5/n)
print("pagerank site6= ",site6/n)
# fin du script
```



1.3 Comprendre le script

En vous aidant du net ou du manuel

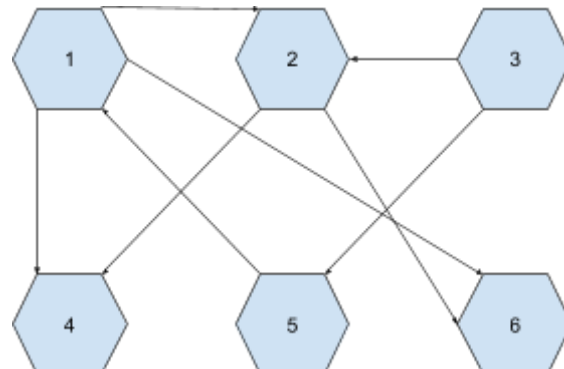
Ligne du script	Expliquez ce que fait ...
<pre>from random import randint</pre>	
<pre>tirage = randint(1,6)</pre>	
<pre>if tirage==1: site1=site1+1</pre>	
<pre>print("Total= ",site1+site2+site3+site4+site5+site6)</pre>	J'obtiens 1000 à chaque fois, pourquoi?
<pre>print("pagerank site4= ",site4/n)</pre>	

2 Seconde étape: sites avec liens (backlinks)

2.1 Second principe

Les sites sont maintenant liés entre eux par des hyperliens. On partira du principe que si un internaute consulte le site 1 alors il suivra tous les liens partant du site 1 vers les sites 2, 4 et 6 comme l'indique le schéma.

Consulter le site 1 revient donc en fait à consulter les sites 1,2,4 et 6 et donc correspond à 4 visites et non une!



2.2 Créer la table des redirections par hyperliens

Le site x	renvoie vers le ou les sites y
site1	site2, site6, site4
site2	
site3	site2, site5
site4	
site5	
site6	

2.3 Créer le second script

Modifier le script initial pour tenir compte de cette nouvelle donnée.

Pour vous aider voici 4 simulations avec 1000 internautes ne jouant qu'une fois. Vous noterez aussi que l'incrémention de n n'est plus à la ligne 9 ;)

```

Thonny - C:\Users\Pierre\Downloads\bbb.py @ 35 : 14
Fichier  Édition  Affichage  Exécuter  Device  Outils  Aide

aaa.py x  bbb.py x
1  # début du script
2  from random import randint
3
4  site1,site2,site3,site4,site5,site6=0,0,0,0,0,0
5  n=0
6
7  for i in range (1000):
8      tirage = randint(1,6)
9
10     if tirage==1:
11         site1=site1+1
12         site2=site2+1
13
14
15
16
17

Console x
>>> %Run bbb.py
328 495 162 516 316 501
Total= 2318
pagerank site1= 0.14150129421915444
pagerank site2= 0.21354616048317515
pagerank site3= 0.06988783433994823
pagerank site4= 0.22260569456427956
pagerank site5= 0.1363244176013805
pagerank site6= 0.21613459879206212

>>> %Run bbb.py
334 488 176 494 350 468
Total= 2310
pagerank site1= 0.1445887445887446
pagerank site2= 0.21125541125541125
pagerank site3= 0.0761904761904762
pagerank site4= 0.21385281385281385
pagerank site5= 0.15151515151515152
pagerank site6= 0.2025974025974026

>>> %Run bbb.py
334 470 154 506 319 491
Total= 2274
pagerank site1= 0.1468777484608619
pagerank site2= 0.20668425681618294
pagerank site3= 0.06772207563764292
pagerank site4= 0.22251539138082674
pagerank site5= 0.14028144239226034
pagerank site6= 0.21591908531222515

>>>
    
```

2.4 Répondre aux questions

Questions	Votre réponse
Pourquoi le total des visites est maintenant aléatoire?	
Ai-je vraiment besoin du n ou puis-je le remplacer par Total?	
Le site 3 est peu visité, est ce normal?	
Les sites 2, 4 et 6 semblent avoir le même nombre de visites, pourquoi?	Parce qu'ils sont pairs ? :))) NON c'est un blague! évidemment que cela n'est pas la réponse!!!

2.5 Votre script

Coller votre script ici

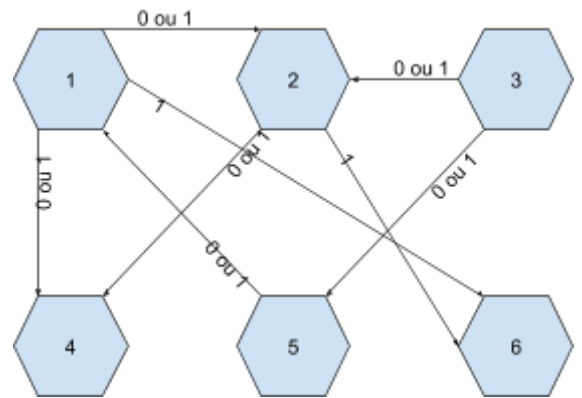
3 Troisième étape: L'influenceur

3.1 Troisième principe

Nous supposons maintenant que le site 6 est géré par un influenceur. Pour traduire cette influence nous dirons par exemple que si un internaute consulte le site 1 alors il consultera aussi le site 6 mais il ira (ou pas) voir les sites 2 et 4.

Cette condition peut se résumer par :

```
if tirage==1:
    site1=site1+1
    site2=site2+randint(0,1)
    site4=site4+randint(0,1)
    site6=site6+1
```



3.2 Mes simulations

Voici quelques simulations

Pour 1 000 internautes	<pre>>>> %Run aaa.py 264 345 152 327 264 502 Total= 1854 pagerank site1= 0.1423948220064725 pagerank site2= 0.18608414239482202 pagerank site3= 0.08198489751887811 pagerank site4= 0.17637540453074432 pagerank site5= 0.1423948220064725 pagerank site6= 0.2707659115426106</pre>
Pour 10 000 internautes	<pre>>>> %Run aaa.py 2531 3357 1664 3363 2488 4989 Total= 18392 pagerank site1= 0.1376141800782949 pagerank site2= 0.18252501087429318 pagerank site3= 0.09047411918225315 pagerank site4= 0.1828512396694215 pagerank site5= 0.13527620704654197 pagerank site6= 0.2712592431491953</pre>
Pour 1 000 000 internautes	<pre>>>> %Run aaa.py 250116 334204 166879 333798 249684 500275 Total= 1834956 pagerank site1= 0.13630626565432633 pagerank site2= 0.18213188763109306 pagerank site3= 0.09094441501594588 pagerank site4= 0.18191062891971252 pagerank site5= 0.13607083766586228 pagerank site6= 0.27263596511305993</pre>

3.3 Votre script

Coller votre script ici

3.4 Répondre aux questions

Questions	Votre réponse
Le pagerank du site 6 tendra-t-il vers 1 si le nombre d'internautes tend vers l'infini?	
YT propose aux internautes des vidéos en fonction de l'historique des vues, finiront-ils par ne voir qu'un seul type de vidéos?	
Pour votre culture	https://www.vice.com/fr/article/zmp4gx/jai-passe-14-heures-daffilee-sur-youtube-pour-etudier-son-algorithme