

Thématique WEB

Le moteur de recherche

Maxime Fourny

Mercredi 5 décembre 2018

plan

- 1 Position du problème
- 2 Comment ça marche
- 3 Et avec les élèves
- 4 Retour sur le modèle

Hiérarchisation de pages web selon une thématique

Hiérarchisation de pages web selon une thématique

- Appel à une sommité

Hiérarchisation de pages web selon une thématique

- Appel à une sommité
- Appel au vote des internautes

Cahier des charges

Cahier des charges

- Exploitable dans tous les domaines

Cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés

Cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données

Cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données
- Prise en compte de l'évolution des données

Cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données
- Prise en compte de l'évolution des données
- Automatisable (donc programmable)

Cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données
- Prise en compte de l'évolution des données
- Automatisable (donc programmable)
- Efficace

plan

- 1 Position du problème
- 2 **Comment ça marche**
- 3 Et avec les élèves
- 4 Retour sur le modèle

Principes généraux

Considérons un certain sujet dont seulement 4 sites l'évoque sur internet.

Principes généraux

Considérons un certain sujet dont seulement 4 sites l'évoque sur internet.

Principes généraux

Considérons un certain sujet dont seulement 4 sites l'évoque sur internet.

- Attribution d'un score ("PageRank") entre 0 (peu pertinent) et 1 (très pertinent) à chaque page

Principes généraux

Considérons un certain sujet dont seulement 4 sites l'évoque sur internet.

- Attribution d'un score ("PageRank") entre 0 (peu pertinent) et 1 (très pertinent) à chaque page
- Affichage par ordre décroissant du score

Principes généraux

Considérons un certain sujet dont seulement 4 sites l'évoque sur internet.

- Attribution d'un score ("PageRank") entre 0 (peu pertinent) et 1 (très pertinent) à chaque page
- Affichage par ordre décroissant du score
- L'idée à la base du modèle de Larry Page et Sergey Brin tient en deux règles liées au référence :

Principes généraux

Considérons un certain sujet dont seulement 4 sites l'évoque sur internet.

- Attribution d'un score ("PageRank") entre 0 (peu pertinent) et 1 (très pertinent) à chaque page
- Affichage par ordre décroissant du score
- L'idée à la base du modèle de Larry Page et Sergey Brin tient en deux règles liées au référence :
 - **R1** : La "PageRank" doit tenir compte des pages qui font référence dans le domaine ("PageRank" élevé).

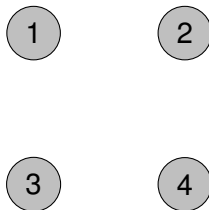
Principes généraux

Considérons un certain sujet dont seulement 4 sites l'évoque sur internet.

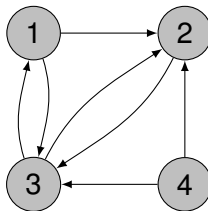
- Attribution d'un score ("PageRank") entre 0 (peu pertinent) et 1 (très pertinent) à chaque page
- Affichage par ordre décroissant du score
- L'idée à la base du modèle de Larry Page et Sergey Brin tient en deux règles liées au référence :
 - **R1** : La "PageRank" doit tenir compte des pages qui font référence dans le domaine ("PageRank" élevé).
 - **R2** : Une référence venant d'une page multipliant les liens doit avoir peu de crédit.

Représentation

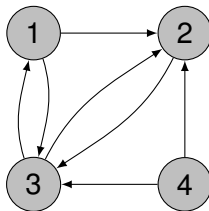
Représentation



Représentation

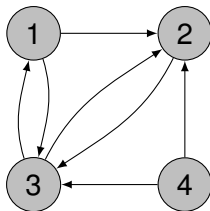


Représentation

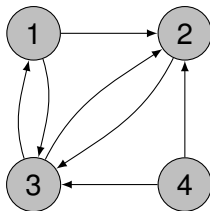


Principe du surfeur aléatoire

Principe du surfeur aléatoire



Principe du surfeur aléatoire

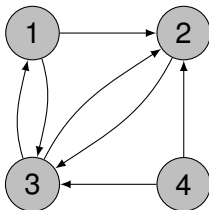


- Matériel : un dé à six faces

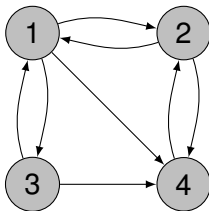
plan

- 1 Position du problème
- 2 Comment ça marche
- 3 Et avec les élèves**
- 4 Retour sur le modèle

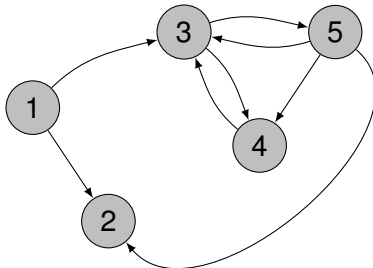
Exercice 1



Exercice 2



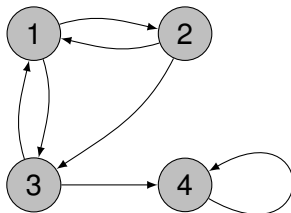
Exercice 3 : premier problème



Exercice 4 : Le gros problème

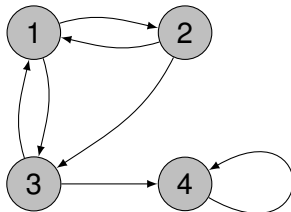
Exercice 4 : Le gros problème

- Le puit :

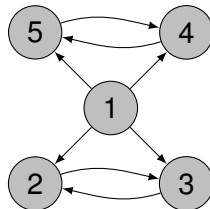


Exercice 4 : Le gros problème

- Le puit :



- La poche web :



Modifications de l'algorithme

Modifications de l'algorithme

- Si une page ne comporte aucun lien vers l'extérieur :
création artificielle de liens vers toutes les autres pages.

Modifications de l'algorithme

- Si une page ne comporte aucun lien vers l'extérieur :
création artificielle de liens vers toutes les autres pages.
- A chaque saut :

Modifications de l'algorithme

- Si une page ne comporte aucun lien vers l'extérieur :
création artificielle de liens vers toutes les autres pages.
- A chaque saut :
 - On continue la promenade aléatoire comme précédemment
avec une probabilité de 0.85 ;

Modifications de l'algorithme

- Si une page ne comporte aucun lien vers l'extérieur : création artificielle de liens vers toutes les autres pages.
- A chaque saut :
 - On continue la promenade aléatoire comme précédemment avec une probabilité de 0.85 ;
 - Avec probabilité de 0.15, on fait un saut aléatoire équiréparti vers n'importe quelle page.

plan

- 1 Position du problème
- 2 Comment ça marche
- 3 Et avec les élèves
- 4 Retour sur le modèle**

Retour sur le cahier des charges

Retour sur le cahier des charges

- Exploitable dans tous les domaines

Retour sur le cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés

Retour sur le cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données

Retour sur le cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données
- Prise en compte de l'évolution des données

Retour sur le cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données
- Prise en compte de l'évolution des données
- Automatisable (donc programmable)

Retour sur le cahier des charges

- Exploitable dans tous les domaines
- Utilisable pour tous les mots clés
- Adaptable à un grand nombre de données
- Prise en compte de l'évolution des données
- Automatisable (donc programmable)
- Efficace

Pourquoi ça marche bien

Pourquoi ça marche bien

- Mathématiques - Chaines de Markov

Pourquoi ça marche bien

- Mathématiques - Chaines de Markov
- Théorème de Peyron Froebenius

Pour aller plus loin

Pour aller plus loin

- On ne s'intéresse aucunement au fond

Pour aller plus loin

- On ne s'intéresse aucunement au fond
- Réflexion sur les règles de calcul d'un score (Yuka)