

L'algorithme *PageRank* de Google

La création de Google

- 1994-1995
 - ◆ Apparition des premiers moteurs de recherche (Lycos, Altavista)
→ résultats pas vraiment bons
 - ◆ Deux étudiants américains Sergey Brin et Larry Page (université de Stanford) se rencontrent



Larry Page

Crédits : Marcin Mycielski, CC BY-SA 4.0



Sergey Brin

Crédits : Steve Juvertson, CC BY 2.0

- ◆ Ils s'associent pour créer un moteur de recherche

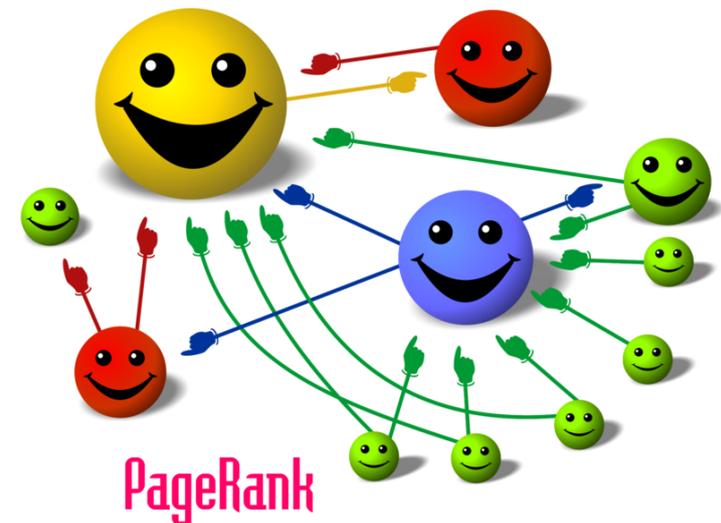
La création de Google

- 1996-1998
 - ◆ Travaillent et développent leurs idées, assez novatrices pour l'époque
 - ◆ Obtiennent des financements
 - ◆ Septembre 1997 : le nom de domaine « google.com » est enregistré
 - ◆ Septembre 1998 : la compagnie *Google Inc.* est créée et s'installe dans un garage loué pour l'occasion



Calculer la popularité d'une page

- L'algorithme PageRank à la base du moteur de recherche Google
 - ◆ PageRank permet de calculer la **popularité** d'une page
 - ◆ A fait la fortune de Google
- Idée de base
 - ◆ Faire confiance aux auteurs des pages Web :
Une page est populaire si elle est pointée par beaucoup de pages
 - ◆ Ne marche pas tout à fait car on peut créer des pages artificielles qui pointe vers notre page



Crédits : Mayhaymate, CC BY-SA 2.5

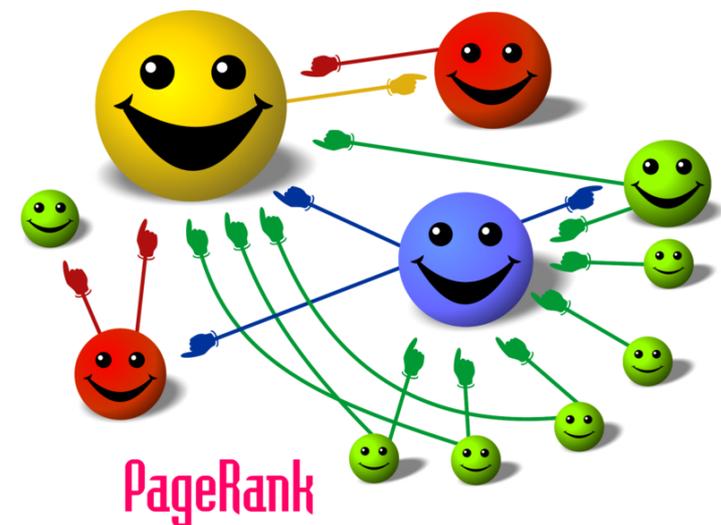
Calculer la popularité d'une page

- Idée de PageRank

- ♦ Donner du poids aux pages populaires :
Une page est populaire si elle est pointée par beaucoup de pages mais qui sont elles-mêmes populaires !

(on obtient une définition circulaire)

- ♦ Nécessite de calculer en même temps la popularité de toutes les pages du web qui dépendent les unes des autres
 - Revient mathématiquement à résoudre une équation gigantesque (autant d'inconnues que de pages Web)
→ on ne sait pas faire !
 - Comment s'en sortir ?
→ s'en remettre au hasard !



Crédits : Mayhaymate, CC BY-SA 2.5

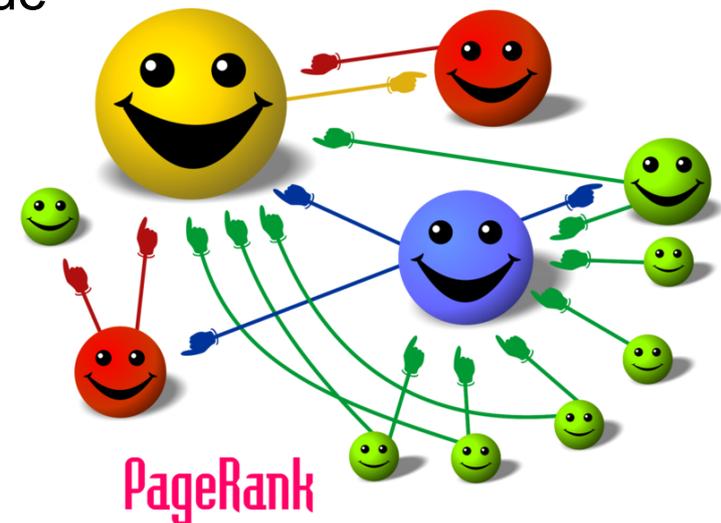
Calculer la popularité d'une page

- Idée de PageRank (suite)

- ◆ On fait des marches aléatoires :

- ◆ On lance en parallèle des millions de marcheurs qui se déplacent au hasard d'une page à l'autre
- ◆ Chaque fois qu'un marcheur passe par une page, il lâche un petit caillou
- ◆ Si on fait cela assez longtemps, la taille du tas de caillou est la bonne solution du problème mathématique

- ◆ Plus une page a été visitée, plus elle est populaire !



Crédits : Mayhaymate, CC BY-SA 2.5

Conclusion

- L'algorithme PageRank permet de calculer la popularité des différentes pages Web.
- Ce n'est qu'*un* des algorithmes utilisés par Google pour classer les résultats d'une requête
- Avez-vous des idées d'autres critères utilisés pour déterminer la pertinence d'une page Web ?

Conclusion

- L'algorithme PageRank permet de calculer la popularité des différentes pages Web.
- Ce n'est qu'*un* des algorithmes utilisés par Google pour classer les résultats d'une requête
- Avez-vous des idées d'autres critères utilisés pour déterminer la pertinence d'une page Web ?
 - ◆ Qualité de la page (taille, vitesse de chargement)
 - ◆ Taux de visite du site
 - ◆ Notoriété, médiatisation du site
 - ◆ Mots clés dans le nom de domaine
 - ◆ Temps de connexion sur la page
 - ◆ ...

Sources

- Podcast de l'émission « Le code a changé » : « Le Français qui a vu naître Google »
<https://www.radiofrance.fr/franceinter/podcasts/le-code-a-change/le-francais-qui-a-vu-naitre-google-4402005>
- Conférence de Gérard Berry « Pourquoi l'informatique met le monde à l'envers » (université de Strasbourg, 2018)
<https://youtu.be/0jqCyosZhFA?t=1805>
- Manuel BORDAS de SNT (collection 3.0)
- Article Wikipédia sur PageRank :
<https://fr.wikipedia.org/wiki/PageRank>