

Dans la peau d'un moteur de recherche : le PageRank

Mise à jour du 20 octobre 2017

Rémi Bachelet

La dernière version de ce cours est ici : [calcul du PageRank](#).

Cette formation est également enregistrée [en vidéo](#)

Cours distribué sous licence **Creative Commons**,
selon les conditions suivantes :



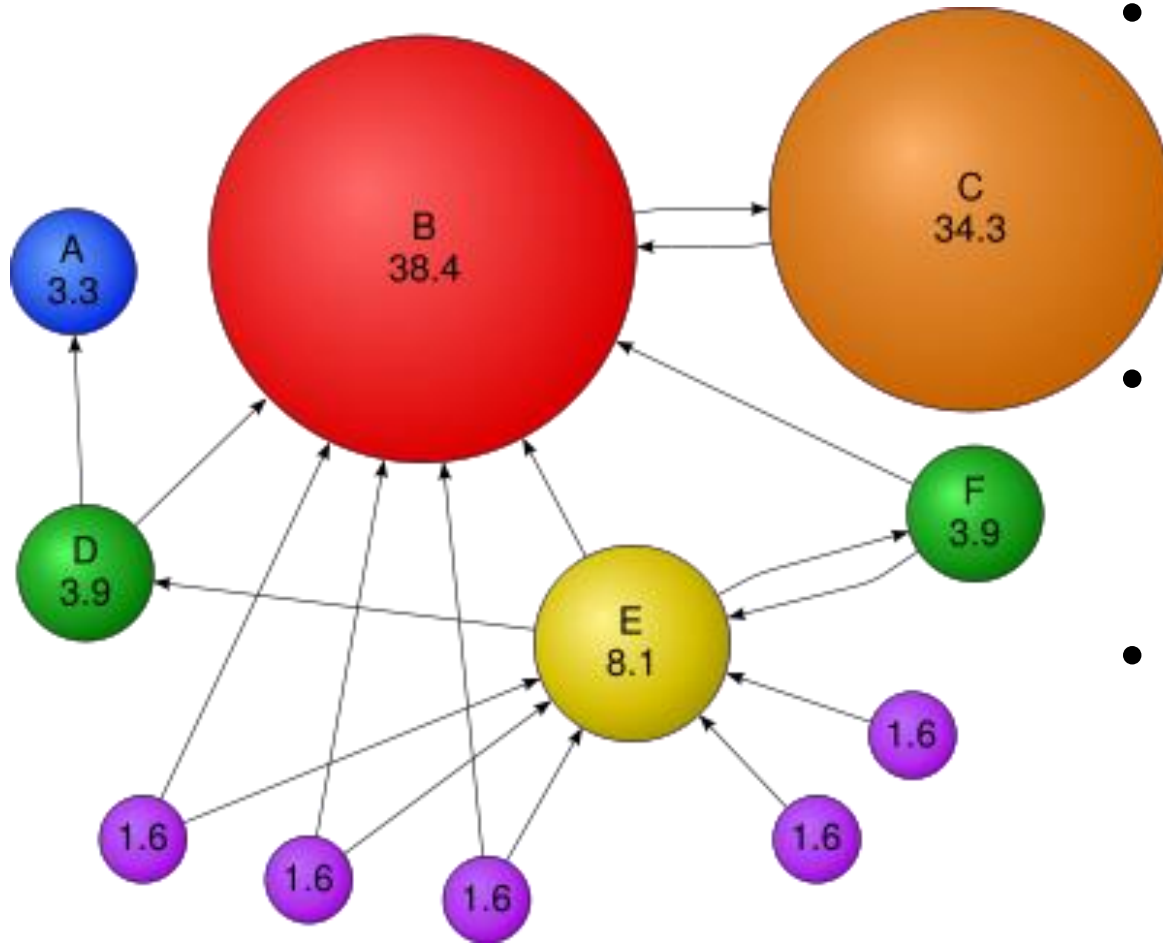
Les algorithmes de classement des pages

1. Le PageRank

- Modalités de calcul
- Fiabilité : Le TrustRank
- Réponse à une requête : Le SERP Rank

Qu'avez-vous retenu ? Quiz rush

Le PageRank



- « Vote » d'une page pour une autre
 - PR (C) > PR (E), même avec moins de liens !
- Échelle logarithmique 0-10
 - Avoir un PR de 3-4 c'est déjà beaucoup
- Un processus de calcul récursif
 - Pour éviter que le PR $\rightarrow \infty$, il faut un amortissement = *Damping factor* (typ. 85% - ici il est de 90%).

L'algorithme du PR : un secret bien gardé

Un nombre important de facteurs est pris en compte dans le PageRank.

- Leur nature et leur pondération sont secrets pour limiter les chances de manipulations (et la concurrence des autres moteurs de recherche...).
- Le terme "PageRank" est une marque déposée et a été l'objet de brevets, à commencer par ([U.S. Patent 6,285,999](#)). Le brevet appartient à Stanford University et Google en a l'usage exclusif, mais l'algorithme a beaucoup évolué depuis le dépôt en 98.
- Beaucoup de spéculations sur ce sujet, voyons quelques-uns des paramètres connus...

Quelques-uns des 200 paramètres du ranking

1. Sur la page (« *onpage* »)
 - Ancienneté / Fréquence d'actualisation
 - Texte = visible sur la page / Code = Meta tags = non visibles sur la page
2. Sur le site (« *onsite* »)
 - Lien internes, arborescence, fil d'ariane (« *Breadcrumbs* »)
 - Paramétrage sur Google Console (Sitemap..)
3. Hors du site (« *offsite* »)
 - Liens entrants en (petite) partie visibles via une recherche Google [link:http://fr.wikipedia.org](http://fr.wikipedia.org)
 - Leur PageRank, Âge, TrustRank de la page
 - Social bookmarking, tweets...

Un débat : Google utilise t-il les données qu'il stocke sur le comportement des internautes pour le calcul du PageRank ?

- Temps passé sur le site, statistiques renvoyées par la barre d'outil google, citations d'URL dans gmail, requêtes avec l'URL du site, marque-pages Google, âge/sexe/localisation des internautes, leurs recherches précédentes *les licences de ces services précisent souvent que non..*

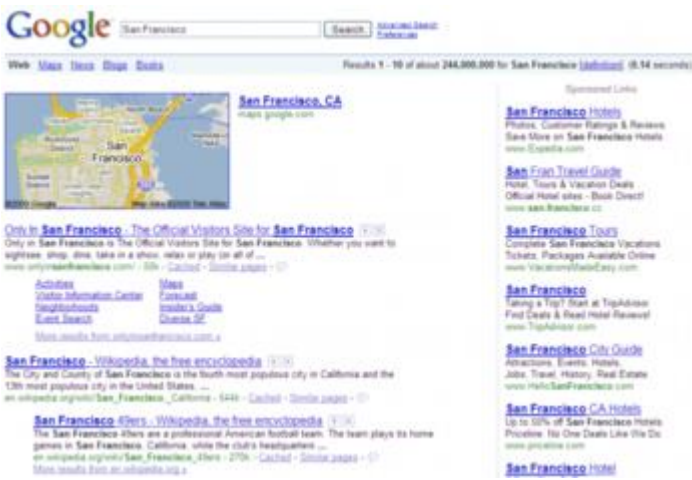
Le PageRank c'est fini ?



- Les mécanismes qui comptent dans l'algorithme actuel ***Hummingbird*** sont nombreux et variés
- Un « Colibri » qui s'alimente à diverses sources selon les situations
 - [RankBrain](#), basé sur l'apprentissage statistique (*machine learning*)
 - Panda : selon la qualité globale d'un site
 - Penguin : anti « black hat »
 - Pigeon : recherches locales,
 - Top Heavy : pages contenant trop de pubs
 - Mobile Friendly : Mobilegeddon
 - Pirate : copyright
 - .. et toujours PageRank : liens entrants

Le TrustRank

- Méthode semi-automatique pour détecter les pages de spam = classification "spam ou pas spam" (*Trust* = confiance - Le terme TrustRank vient de Yahoo!).
 - Principe : une page « propre » ne propose pas de liens vers des pages de spam
1. Amorçage : établir une liste de pages « propres » de référence
 - Après une analyse « humaine ».
 - On n'a pas forcément besoin d'une grande liste (p.e 200 sites).
 2. Suivi récursif des liens de la liste d'amorçage
 3. Degré de confiance que l'on peut attribuer à la page : un indice
 - Plus les liens sont forts avec des pages de référence, plus leur degré de confiance est élevé
 - C'est le TrustRank (ou TR), indice entre 0 (=spam) et 1 (=page de référence)
- Le TrustRank peut être utilisé :
 - pour filtrer l'index d'un moteur de recherche,
 - pour classer les résultats d'une recherche.



Le SERP Rank

C'est l'ordre de présentation des liens lorsque l'on entre des mots-clés dans un moteur de recherche

- La page de résultats présente une liste ordonnée de liens vers des pages/images/vidéos, associés à des textes courts (*snippets*)
- Le SERP Rank est fonction du PageRank, mais aussi de facteurs liés aux mots-clés.
 - Voir le chapitre 6 sur les mots-clés et leur mise en valeur
 - SERP = *Search Engine Results Page*

News results for **paris hilton**



[Lindsay Lohan and Paris Hilton pay tribute to Johnson & Johnson ...](#) - 6 days ago
Friends Lindsay Lohan and **Paris Hilton** also paid tribute on the site. **Paris** wrote: "I feel so sad, my heart is in pain. Just lost a longtime friend. ..."
[Mirror.co.uk](#) - [3164 related articles >](#)
[Paris Hilton to tie the knot soon? - Times of India](#) - [221 related articles >](#)
[Paris Hilton Plays Paparazzi at Aunt's Birthday](#) - [E! Online \(blog\)](#) - [25 related articles >](#)

News Results

Paris Hilton - Wikipedia, the free encyclopedia

Paris Whitney Hilton (born February 17, 1981) is an American socialite, heiress, media personality, model, singer, author, fashion designer and actress. ...
en.wikipedia.org/wiki/Paris_Hilton - [Cached](#) - [🗨](#) [🔗](#) [✕](#)

Web Results

Paris Hilton | The Official Website

[ParisHilton.com](#) **Paris Hilton**, Nicky Hilton Fashion, Pictures, Apparel, Jewellery, Film, and Fun.
www.parishilton.com/ - [Cached](#) - [Similar](#) - [🗨](#) [🔗](#) [✕](#)

Paris Hilton Zone | **Paris Hilton** Pictures, Pics, Photos

8000+ new **Paris Hilton** pictures, Paris wallpaper, lyrics, audio, video, daily Paris pics & news.
[Video](#) - [Wallpaper](#) - [7000+ Paris Pics & Photos](#) - [TV / Films](#)
www.parishiltonzone.com/ - [Cached](#) - [Similar](#) - [🗨](#) [🔗](#) [✕](#)

Image results for **paris hilton** - [Report images](#)



Image Results

Video results for **paris hilton**



[Paris Hilton Responds to McCain Ad](#)
1 min 51 sec - 7 Sep 2008
sjl.funnyordie.com



[Paris Hilton Carls Jr Burger Ad Extended ...](#)
1 min 1 sec - 8 Feb 2008
www.youtube.com

Video Results

Paris Hilton

Socialite **Paris Hilton** was born on February 17, 1981 in New York City... Visit IMDb for Photos, Filmography, Discussions, Bio, News, Awards, Agent, ...
www.imdb.com/name/nm0385296/ - [Cached](#) - [🗨](#) [🔗](#) [✕](#)

Google Universal Search

- Depuis mai 2007, une recherche sur Google propose sur sa **search engine results page (SERP)** des résultats mixant :
 - Certes
 - des Pages web
 - Des liens sponsorisés
 - Mais aussi
 - Actualités,
 - Images,
 - Livres
 - Cartes
 - Blogs
 - Vidéos
 - Définitions (*define:*)
 - Avec *Universal Search*, des contenus auparavant séparés sont présentés simultanément
 - Résultat :
 - encore plus de candidats à la première page
 - + Des stratégies alternatives pour y figurer
- Une évolution incessante : nouvelles versions régulièrement

Questions ?

- Outils d'analyse <https://rankinity.com>
- EdgeRank de Facebook
- Plus d'informations sur les [lien-retours/backlinks](#)
- Mathématiquement, le *PageRank* est la [probabilité stationnaire d'une chaîne de Markov](#), c'est-à-dire un vecteur de [Perron-Frobenius](#) de la [matrice d'adjacence](#) du graphe du Web^{[1],[2]}

SEO /3 PageRank [Quiz rush] <https://goo.gl/GzNXzV>

Les thèmes et chapitres du cours

1. Origine du SEO, Google ... et ses concurrents
2. La fréquentation d'un site : les fondamentaux
3. Dans la peau d'un moteur de recherche : le PageRank
4. Web Analytics et liens commerciaux
5. Optimisation du référencement
 - “Onpage”
 - “Onsite”
 - “Offsite”
6. Trouver et optimiser les mots-clés
7. Méthodologie de référencement et avenir du SEO