

Grand Orale

Sommaire

Sujet Maths-Nsi

Comment a évolué le web et ses algorithmes tels que Page-Rank ?

Sujet Maths-Nsi

Pourquoi peut-on dire que le Page-Rank a révolutionné le web ?

Histoire du web

Pagerank Histoire

Comment est calculé le score du Google PageRank ?

Conclusion

Histoire du web

Vous utilisez le web tous les jours, mais connaissez l'histoire qui se cache derrière et comment est venue l'idée de créer une sorte de réseau permettant de mettre en relation plusieurs articles. Aujourd'hui je vais vous donner des réponses.

Le mot WEB vient du mot World Wide Web, la Toile mondiale ou la Toile, c'est un système de liens hypertexte public fonctionnant sur Internet, attention à ne pas confondre les deux. Le Web permet de consulter, avec l'aide d'un navigateur, des pages accessibles sur des sites. L'image de la toile d'araignée vient des hyperliens qui lient les pages web entre elles

C'est en 1989 que Tim-Berners-Lee est le principal créateur du web, son idée était de créer un système hypertexte qui une fois distribué sur le réseau informatique les collaborateurs puissent partager les informations au sein du CERN. L'Organisation européenne pour la recherche nucléaire (CERN)

Pagerank Histoire

Le PageRank (ou PR) est un système de **classement** des pages qui est utilisé pour **attribuer à chaque page un score relatif d'importance** et d'autorité en évaluant la **qualité** et la **quantité** de des **liens** externes qu'elle reçoit.

Il s'agit de **l'algorithme de base** du moteur de recherche de **Google** à sa création et la philosophie qui a fait de **Google le moteur de recherche le plus pertinent et numéro 1 mondial**

Les premiers moteurs de recherche tels que Yahoo, Lycos, Altavista offraient des **résultats** aux utilisateurs **qui n'étaient pas toujours très pertinents**, loin de là.

De ce constat, les **2 cofondateurs du moteur de recherche Google**, Sergey Brin et Larry Page, ont conçu **PageRank en 1997** à l'université de Stanford dans le cadre d'un projet de recherche doctorale.

L'idée des deux créateurs a été inspirée par la façon dont les scientifiques évaluent « **l'importance** » des articles scientifiques en examinant le nombre d'autres articles scientifiques y faisant référence. Ils ont repris ce concept et l'ont appliqué au Web en suivant les liens entre les pages Web comme des références, des votes positifs.

Ce concept de Pagerank est ensuite devenu la base du moteur de recherche Google et reste l'une des bases du classement des sites aujourd'hui.

Comment est calculé le score du Google PageRank ?

PageRank fonctionne en **comptant le nombre et la qualité des liens vers une page** pour déterminer une **estimation approximative de l'importance du site web**. L'hypothèse sous-jacente est que les sites internet les plus importants sont susceptibles de recevoir davantage de liens venant d'autres sites, les fameux backlinks ou liens entrants.

Le Google PageRank est calculé par une formule mathématique qui compte **tous les liens vers un site Web comme un vote**. Le PR de chaque page dépend du PR des pages **pointant vers elle**.

Formule 1 $PR(P) = \sum_{i \in in(p)} \frac{PR(i)}{out(i)}$ Le Page rank de la page P est la somme de tous les pages ranks des pages ayant un lien vers P

Sauf que certaine page n'ont pas de lien entrant mais on veut quand même attribuer une certaine importance Ici on se place dans la situation de se trouver sur un page web donnée par quelqu'un qui parcourait le web de façon aléatoire

D : Quand je suis sur un âge donné j'ai une probabilité D de rester sur cette page et de cliquer sur un des liens de la page

1 - D: quitter cette page et aller sur une page aléatoire

Toutes les pages on une proba de $1-p / N$ d'être choisi

$\sum_{i \in in(p)}$ (Somme sur tous les éléments qui ont un lien vers P in(p))

$\frac{PR(i)}{out(i)}$ (Pour chaque Élément i on prend le page rank de i divisé par le nombre de lien sortant

Formule 2 $PR(P) = \frac{1-d}{N} + d * \sum_{i \in in(p)} \frac{PR(i)}{out(i)}$

Exemples

A1, A2, A3, A4..., An sont des pages faisant un lien vers une page de destination appelée B.

PR (Ak) est le PageRank **de la page Ak**.

N (Ak) est le **nombre de liens sortants présents sur la page Ak**.

d est un **facteur d'amortissement** compris entre 0 et 1, fixé en général à 0,85.

Le PageRank de la page B se calcule à partir du PageRank de toutes les pages Ak :

$$PR(B) = (1-d) + d \times (PR[A1]/N[A1] + \dots + PR[An]/N[An])$$

Notez que les PageRanks forment une distribution de probabilité sur les pages web, la somme des PageRanks de toutes les pages Web étant égale à un.

Cela signifie, en termes plus simples, que le PageRank de la page B est calculé en multipliant le PageRank de la page A par 0,85. C'est ce qu'on appelle le facteur d'amortissement.

Si la Page B est ensuite liée à la Page C, celle-ci reçoit 85 % du PageRank de B (72,25 % de celui de A).

Par itération, le calcul s'actualise et se stabilise une fois que les Pagerank de toutes les pages liantes sont calculés.

Le PageRank de la page B dépend donc de 3 facteurs :

- le nombre de pages Ak faisant un lien vers B.
- le Pagerank des pages Ak.
- le nombre de liens que la page Ak fait vers d'autres pages.

Cas particulier :

Si une page Web ne comporte **pas de liens pointant vers elle**, cette page ne commence pas par un PageRank de 0, mais plutôt de **0,15**.

PageRank Algo

Le **théorème de point fixe** est le concept mathématique qui a rendu possible le calcul du *PageRank*

<http://infolab.stanford.edu/~backrub/google.html> site offi

<https://fr.semrush.com/blog/pagerank-google/> 0,85

<https://www.netoffensive.blog/referencement-naturel/premier-sur-google/fonctionnement/algorithmes-google/pagerank/#:~:text=Le%20Google%20PageRank%20est%20calcul%C3%A9,des%20pages%20pointant%20vers%20elle.>

<https://www.bibmath.net/quotidien/index.php?action=affiche&quoi=pagerank>

https://www.pedagogie.ac-nantes.fr/medias/fichier/theme-web-fiche-2-05-pagerank-stefperk_1562840627493-pdf?ID_FICHE=1424110212672&INLINE=FALSE#:~:text=Le%20PageRank%20ou%20PR%20est,popularit%C3%A9%20d'une%20page%20web.