



### TF/IDF (1970)

Le **TF/IDF** est une technique de recherche documentaire qui attribue un **score** à une page. Un score plus élevé est censé représenter une **meilleure pertinence**.

Ce score est proportionnel à la **fréquence** du mot recherché dans le fichier.

On peut abuser de cette méthode de recherche : augmenter la **fréquence** des mots-clés fait augmenter le score de la page.

### PageRank (1998)

Le **PageRank** est un algorithme mis en place par Larry Page, cofondateur de Google, pour proposer les résultats les **plus pertinents**.

Son but est de donner une note à un site qui illustrera sa **popularité**.

Celle-ci est calculée en fonction du nombre de liens qui pointent vers la page en question et de la **qualité** des pages qui les contiennent.

### Filter Bubble (2010)

Internet a été vu comme un moyen **d'accéder totalement** à l'information, mais depuis peu les moteurs de recherche personnalisent les résultats en fonction de l'utilisateur.

Ceci est source d'une **bulle de filtres** avec un univers personnel d'informations. Les moteurs de recherche nous montrent donc ce que nous voulons voir, pas nécessairement ce que nous avons besoin de voir.

1970

1980

1990



YAHOO!



2000



2010



### SEO : Optimisation des Moteurs de Recherche (1994)

Google a rédigé un **guide d'optimisation** qui liste les méthodes pour qu'un site soit bien référencé (réfléchir au titre, aux descriptions du contenu, des images...).

Il existe néanmoins **d'autres méthodes** moins honnêtes comme dupliquer le contenu, notamment les mots-clés, afficher une description et un titre ne correspondant pas à la page, ou encore masquer des liens ou des textes.

### Schema.org (2011)

**Schema.org** est une organisation ayant pour but la promotion de schémas servant à structurer les données sur internet, notamment les sites web. Ces schémas constituent une **liste d'attributs** à rajouter au sein même des balises HTML englobant le contenu.

Les moteurs de recherche peuvent ensuite utiliser ces données pour chercher du **sens** en plus du seul texte brut, de manière à améliorer les résultats ou proposer directement l'information recherchée (ce que fait le **KnowledgeGraph** de Google).