

PageRank

Da Wikipedia, l'enciclopedia libera.

Il **PageRank** è un algoritmo di analisi che assegna un peso numerico ad ogni elemento di un collegamento ipertestuale d'un insieme di documenti, come ad esempio il World Wide Web, con lo scopo di quantificare la sua importanza relativa all'interno della serie. L'algoritmo può essere applicato a tutti gli insiemi di oggetti collegati da citazioni e riferimenti reciproci. Il peso numerico assegnato ad un dato elemento E è chiamato anche "il PageRank di E", siglato in PR (E). L'algoritmo di PageRank è stato brevettato (brevetto US 6285999 (<http://v3.espacenet.com/publicationDetails/biblio?CC=US&NR=6285999&KC=&FT=E>)) dalla Stanford University; è inoltre un termine ormai entrato di fatto nel lessico dei fruitori dei servizi offerti dai motori di ricerca. Il nome PageRank è un marchio di Google.



Schematizzazione del sistema
PageRank

Indice

- 1 Elementi generali
- 2 Visualizzazione del grado di popolarità di un sito
- 3 Formula semplificata
- 4 Voci correlate
- 5 Altri progetti
- 6 Collegamenti esterni
- 7 Note

Elementi generali

Letteralmente traducibile come *rango* di una pagina web (ma anche un gioco di parole collegato al nome di uno dei suoi inventori, Larry Page), il *pagerank* è facilmente riconducibile al concetto di popolarità tipico delle relazioni sociali umane, ed indica, o si ripromette di indicare, le pagine o i siti di maggiore rilevanza in relazione ai termini ricercati. Gli algoritmi che rendono possibile l'indicizzazione del materiale presente in rete utilizzano anche il grado di popolarità di una pagina web per definirne la posizione nei risultati di ricerca.

Questo metodo può esser descritto come analogo ad una elezione nella quale ha diritto al voto chi può pubblicare una pagina web, e il voto viene espresso attraverso i collegamenti in essa presenti. I voti non hanno tutti lo stesso peso: le pagine web più popolari esprimeranno, coi propri *link*, voti di valore maggiore.

L'interpretazione e la definizione della popolarità di un sito non sono però legate soltanto a queste *votazioni*, ma tengono conto anche della pertinenza del contenuto di una pagina, nonché delle pagine correlate, con i termini ed i criteri della ricerca effettuata. Altro importante elemento che lega un sito alla sua popolarità è relativo alla diffusione, alla popolarità dell'argomento trattato in esso. Per argomenti poco

richiesti i siti raggiungono facilmente le prime posizioni nelle ricerche, ma altrettanto verosimilmente posseggono e mantengono un page rank che potrebbe essere bassissimo. Tutto questo permette, o perlomeno ha lo scopo, di attuare un controllo incrociato che garantisca la validità dei risultati di ricerca.

Visualizzazione del grado di popolarità di un sito

Per ottenere informazioni riguardo, ad esempio, alla *notorietà* di una pagina web in Google è possibile installare la Google Toolbar nel browser di navigazione oppure si possono utilizzare servizi proposti da terze parti in grado di determinare tale informazione ed implementabili su siti internet. Per aumentare il page rank si possono sfruttare dei piccoli accorgimenti. Uno dei più usati è cercare di attrarre l'attenzione di utenti che si interessino ad argomenti correlati al nostro sito o blog. Lo scopo è conquistare gli utenti veramente interessati a seguire i nostri post. Tale accorgimento serve per scavalcare una delle regole di Google: i link che possono puntare a un sito devono essere "naturali", ovvero non richiesti o generati da scambio di link o a pagamento. Se Google si accorge che un sito fa richieste che non rispettano le regole del motore di ricerca, lo penalizza immediatamente, arrivando anche a bannare il proprietario che vada contro le regole di condotta sottoscritte al momento dell'iscrizione al servizio del motore di ricerca ^[1]. Per cercare di ottenere link naturali è necessario proporre argomenti sempre aggiornati, non copiati, con approfondimenti interessanti per i fruitori del servizio web.

Formula semplificata

L'algoritmo completo per il calcolo del PageRank fa ricorso all'uso della teoria dei processi di Markov ed è classificato nella categoria degli algoritmi di Link Analysis Ranking. Dalla formula inizialmente sviluppata dai fondatori di Google, Sergey Brin e Larry Page, è possibile comprendere come il PageRank viene distribuito tra le pagine:

$$PR[A] = \frac{1-d}{N} + d \left(\sum_{k=1}^n \frac{PR[P_k]}{C[P_k]} \right)$$

Dove:

- $PR[A]$ è il valore di PageRank della pagina A che vogliamo calcolare.
- N è il numero totale di pagine note.
- n è il numero di pagine che contengono almeno un link verso A. P_k rappresenta ognuna di tali pagine.
- $PR[P_k]$ sono i valori di PageRank di ogni pagina P_k .
- $C[P_k]$ sono il numero complessivo di link contenuti nella pagina che offre il link.
- d (*damping factor*) è un fattore deciso da Google e che nella documentazione originale assume valore 0,85. Può essere aggiustato da Google per decidere la percentuale di PageRank che deve transitare da una pagina all'altra e il valore di PageRank minimo attribuito ad ogni pagina in archivio.

Dalla formula si nota quindi che all'aumentare del numero di link complessivi dei siti che puntano ad A il PageRank aumenta.

Voci correlate

- Googlebombing
- Metodo delle potenze: metodo iterativo per il calcolo di autovettori utilizzato nell'algoritmo Page

rank

- Processo markoviano
- TrustRank

Altri progetti

-  **Commons** ([//commons.wikimedia.org/wiki/Pagina_principale?uselang=it](https://commons.wikimedia.org/wiki/Pagina_principale?uselang=it)) contiene immagini o altri file su **PageRank** ([//commons.wikimedia.org/wiki/Category:PageRank?uselang=it](https://commons.wikimedia.org/wiki/Category:PageRank?uselang=it))

Collegamenti esterni

- Tecnologia *PageRank* di Google (<http://www.google.com/corporate/tech.html>)
- (EN) Sergey Brin, Larry Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine* (<http://dbpubs.stanford.edu/pub/1998-8>), 1998. URL consultato il 26 febbraio 2009.

Note

- ↑ Ciò porrebbe diverse problematiche inerenti alla fruizione dei servizi commerciali quali ad esempio la pubblicità del proprietario del sito che potrebbe rischiare anche di non rientrare nel circuito di Google.

 **Portale Google**

 **Portale Web**

Categorie: Ricerche di Google | Motori di ricerca | Siti web | [altre]

-
- Questa pagina è stata modificata per l'ultima volta il 28 gen 2014 alle 03:07.
 - Il testo è disponibile secondo la licenza Creative Commons Attribuzione-Condividi allo stesso modo; possono applicarsi condizioni ulteriori. Vedi le Condizioni d'uso per i dettagli. Wikipedia® è un marchio registrato della Wikimedia Foundation, Inc.