

Il PageRank di Education 2.0

L'approfondimento del significato di PageRank, come indice di riferimento della visibilità e attendibilità del contenuto proposto da un sito web, nasce da una cordiale conversazione tra chi scrive e l'ex Ministro della Pubblica Istruzione Luigi Berlinguer, in una pausa dell'interessantissimo secondo convegno di Education 2.0. Il concetto a cui si ispira il calcolo del PageRank per un sito web deriva da considerazioni di tipo accademico, basate sul quando, in ambito universitario, una pubblicazione, una ricerca, un articolo, acquista importanza e notorietà. La risposta a questa domanda è immediata e si fonda sul fatto che tra i testi accademici, quelli più prestigiosi si misurano per quante più citazioni accumulano. Quindi non conta il numero di copie vendute, ma il numero di colleghi che lo citano e lo prendono come esempio e come modello. Sono i "pari" che conferiscono a un testo scientifico uno status di importanza e di notorietà. Per quanto detto, i webmaster attribuiscono importanza a un sito web, mettendo un link esterno, con il significato di valutare una risorsa web considerata di buona qualità. Inoltre un webmaster, consigliando (sotto forma di link) una buona pagina web ai propri utenti, aumenta anche le visite al proprio sito, in altre parole realizza il proprio interesse di visibilità nella rete. Quindi, se un webmaster ha un obiettivo personale a inserire un link esterno, si presume che sia convinto della buona qualità del sito linkato. Questo elementare concetto di valutazione ha dato vita all'enorme successo di Google, che appunto si è sempre basato sulla considerazione e sul peso dei link che puntano a un determinato documento.

L'algoritmo completo per il calcolo del PageRank, inizialmente sviluppato dai fondatori di Google, Sergey Brin e Larry Page, è classificato nella categoria degli algoritmi di Link Analysis Ranking, e lo sviluppo della sua formula è riportato di seguito:

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

Dove:

- 1) PR[A] è il valore di PageRank della pagina A che vogliamo calcolare.
- 2) n è il numero di pagine che contengono almeno un link verso A. P_k rappresenta ognuna di tali pagine.
- 3) PR[P_k] sono i valori di PageRank di ogni pagina P_k.
- 4) C[P_k] sono il numero complessivo di link contenuti nella pagina che offre il link.
- 5) 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 a ogni pagina in archivio.

Per fare un esempio numerico della formula del PageRank, immaginiamo che la

rivista on line Education 2.0, abbia 12 pagine esterne che la menzionino, ognuna delle quali avente un link riferito a essa, su un totale di 8 link presenti per pagina. Di queste 12 pagine tre hanno PageRank pari a 3, quattro pari a 4 , altre quattro pari a 5 e una pari a 6. Il PageRank di Education 2.0 sarà: $PR(\text{Education 2.0}) = (1 - 0.85) + 0.85 (3+3+3+4+4+4+4+5+5+5+5+6)/8 = 0.15 + 39.95/8 = 5$

Quindi il sito Education 2.0 avrà un PageRank di 5, perché molti siti con PageRank uguale la citano. In altre parole ciò vuol dire che (a parità di altri elementi) un link ricevuto da una pagina che possiede un PageRank alto "vale" più di un link ricevuto da una pagina dal valore di PageRank basso e contribuisce maggiormente a far aumentare il valore di PageRank della pagina che riceve il link.

Oltre il PageRank, i motori di ricerca utilizzano la Link Popularity (LP) per attribuire la posizione a una pagina web, e questa variabile coincide col numero complessivo di link che puntano a quella pagina. Maggiore è il numero dei link che puntano a una pagina e migliore sarà la posizione di quella stessa pagina nei risultati delle ricerche. Ne consegue che i motori di ricerca che usano (parzialmente) la LP, daranno una preferenza ai siti che ricevono un numero di link maggiore rispetto agli altri. Questo criterio, però, presenta alcune criticità, per esempio penalizza molto i piccoli siti appena costituiti, che necessitano di molto tempo prima di ottenere una buona quantità di link da altri siti. Quindi possiamo dire che il PageRank al contrario della Link Popularity, invece di tener conto solo della quantità dei link ricevuti, prende in considerazione anche il valore di PageRank delle pagine che offrono i link.

Da quanto detto Education 2.0 può vantare con merito un PageRank 5 (tra i valori più alti per riviste on line italiane che si occupano di scuola), nonostante i soli due anni di vita, perché riesce a ottenere molti link da siti web qualificati e affidabili nei contenuti, ponendo le basi, nel breve e medio periodo, a una progressione di visibilità nel web del tutto invidiabile.

Aldo Domenico Ficara