The Second International Conference "Problems of Cybernetics and Informatics" September 10-12, 2008, Baku, Azerbaijan. Section #1 "Information and Communication Technologies" www.pci2008.science.az/1/29.pdf

APPLICATION OF PAGE RANK ALGORITHM IN EVALUATION OF SCIENTIFIC ARTICLES

Rahile Hasanova

Institute of Information Technology of ANAS, Baku, Azerbaijan
rahasanova@gmail.com

Larry Page and Sergey Brin first investigated evaluation of quality indicators of Internet pages, called Page Rank at Stanford University. Having started in 1995, this project led to development of the functional system, which had successfully stood a test within Google system. A program for automatic finding of the arbitrary Internet site has been developed and at the present by writing the address of any page of our interest in a small window, we can calculate frequency of usage of the same page in the Internet by applying 10 points system.

Algorithm of the web page evaluation – Page Rank is thought to be applied to find Article Rank of the article. Let us first consider PageRank algorithm:

$$PR(p_i)=\frac{1-d}{N}+d\sum_{p_j\in B(p_i)}\frac{PR(p_j)}{O(p_j)}$$

where,

$p_1, p_2, ..., p_N$ – pages viewed,

$B(p_i)$ – number of the pages referring to page p_i,

$O(p_j)$ – p_j number of the references generated from the page,

N – general number of the pages

Values of PageRank are elements of main special vector of modificated neighborhood matrix. Special vector

$$R=\begin{bmatrix} PR(p_1) \\ PR(p_2) \\ \vdots \\ PR(p_N) \end{bmatrix}$$

Is a solution of the following equation

$$R=\begin{bmatrix} (1-d)/N \\ (1-d)/N \\ \vdots \\ (1-d)/N \end{bmatrix}+d\begin{bmatrix} l(p_1,p_1) & l(p_1,p_2) & \cdots & l(p_1,p_N) \\ l(p_2,p_1) & \ddots & & \\ \vdots & & l(p_i,p_j) & \\ l(p_N,p_1) & & & l(p_N,p_N) \end{bmatrix}R$$

$l(p_i,p_j)$ function equals 0 when there is no reference from page p_j to page p_i and is normalized as follows:

$$\sum_{i=1}^{N} l(p_i,p_j)=1,\ \forall i$$

that's elements of each column equals 1.

PageRank special vector's values can be quickly approximated and it yields good results.

Let us now apply the same algorithm to valuation of the scientific articles.

In order to do it let us use the same PageRank algorithm for any u article to find its Article Rank value:

$$AR(u) = \frac{1-d}{n} + d \sum_{v \in B(u)} \frac{AR(v)}{N_v} \cdot N_{vu}$$

where,

$B(u)$ – is collection of articles referring to article u;

N_v – is total number of references generated from article v;

N_{vu} – is a number of references made from article v to article u;

d – is coefficient of extinction, serves for accumulation of the process.

The value of u article's Article Rank is dependent on each of v articles which refer to $B(u)$ majority.

In order to increase objectiveness of the evaluation it is intended to multiply the value of each article's Article Rank with the impact-factor of the journal where the article is published. Our model shapes like this:

$$AR(u) = \frac{1-d}{n} + d \sum_{v \in B(u)} \frac{AR(v) \cdot N_{vu}}{N_v} IF(v)$$

where,

$IF(v)$ – is an impact-factor of the journal where v article is published..

Literature

1. Brin S., Page L. The anatomy of a large-scale hypertextual web search engine // Computer Networks ISDN Systems, 1998, vol. №№ 1-7, pp. 107-117.
2. Ma N., Guan J., Zhao Y. Bringing PageRank to the citation analysis // Information Processing and Management, 2008, vol.44, № 2, pp. 800-810.
3. Pislyakov V.V. Why create national indices of citation? http://inf.by/library/1795/