PageRank: Quantitative Model of Interaction Information Retrieval

Sandor Dominich

Department of Computer Science, University of Veszprém
8200 Veszprém, Egyetem u. 10, Hungary,
Tel.: +36 88 422022/4711
Email: dominich@dcs.vein.hu

ABSTRACT
The PageRank method is used by the Google Web search engine in ranking Web pages, whose PageRank values can be conceived as the steady state distribution of a Markov chain. The Interaction Information Retrieval (I^2R) method is used by the I^2R non-classical information retrieval (IR) paradigm, which represents a connectionist approach based on dynamic systems. In this paper, it is shown that the principles and models used by the PageRank and I^2R methods are equivalent, and that PageRank may be conceived as a particular connectionist dynamic system that is looking for equilibrium in the state space.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *retrieval models*. G.3 [**Probability and Statistics**]: *Markov processes*. I.2.6 [**Artificial Intelligence**]: Learning – *connectionism and neural nets.*

General Terms
Algorithms, Performance, Theory

Keywords
World Wide Web, Retrieval, Google, PageRank, Interaction Information Retrieval, Network Equation, Equilibrium

1. INTRODUCTION
The PageRank method [4] is an important component of the Google Web Information Retrieval (IR) engine [17], and makes it possible the calculation of a priori importance measures for Web pages. The measures are computed offline, and are independent of the search query. At query time, the measures are combined with query-specific scores to obtain a ranking of Web pages. The PageRank method uses the metaphor of an easily bored Web surfer. The PageRank value of a page is conceived as being the probability that the surfer reaches that page by following forward links. The PageRank values form a probability distribution over the Web. This stochastic approach is treated within the wider framework of Markov chains [2], many interesting properties are established and discussed, and techniques are suggested to enhance the PageRank computation. Also, the view based on link analysis was used in enhancing the indexing of Web pages [21]. In the present paper, a different view on PageRank is proposed. It is shown that it can be conceived as a particular case of the Interaction Information Retrieval (I^2R) method [11]. Thus, PageRank can be looked at from a connectionist (dynamic systems) standpoint, too.

2. PAGERANK
The Google search engine exploits the citation graph of the crawled portion of the publicly accessible World Wide Web, and calculates a measure of the relative importance of Web pages using a stochastic process view.

2.1. Place of PageRank in Google's Retrieval and Ranking
The retrieval and ranking of Web pages follows a usual IR scenario, and is performed in several steps which seem to include parts of classical techniques (e.g., Boolean model at point (a) below to locate Web pages containing query terms, vector space model at point (b) below in that numeric vectors are defined and their dot product is taken), and a technique called PageRank as follows: (a) Find the Web pages containing the query terms. (b) Compute a relative importance of Web pages. (c) Rank the Web pages according to their relative importance. The relative importance of Web pages is calculated taking into account several factors such as: (i) 'on page factors', i.e., terms occurring in title, anchor, body, proximity of terms, (ii) appearance of terms: small font, large font, colour, (iii) frequency of occurrence of terms, (iv) PageRank values, (v) other factors.

2.2. The PageRank Method
The PageRank method is considered to be an important factor used by Google in computing the relative importance of Web pages. The PageRank value of a Web page depends on the PageRank values of pages pointing to it and on the number of links going out of these pages.

2.2.1. The Principle
The starting point for the principle of PageRank is citation analysis, which is concerned with the study of citation in the scientific literature [15]. The underlying idea, which is well-known (and has a tremendous literature), of citation analysis reads as follows: citation counts are a measure of importance [16]. This citation idea was used in [6] for Web IR. In the PageRank method [27], this idea is refined in that citation counts are not absolute values anymore, rather relative ones and mutually dependent (as will be seen below). The principle on which PageRank is based can thus be referred to as an extended citation principle, and can be formulated as follows: A Web page's importance is determined by the importance of Web pages linking to it.

2.2.2. The Model
In order to apply the extended citation principle in practice, an appropriate model of a system of Web pages has been constructed as follows. Let (Figure 1)

(i) $\Omega = \{W_1, W_2, \ldots, W_i, \ldots, W_N\}$ denote a set of Web pages under focus,

(ii) $\Phi_i = \{W_k \mid k = 1, \ldots, n_i\}$ denote the set of Web pages W_i points to, $\Phi_i \subseteq \Omega$,

(iii) $B_i = \{W_j \mid j = 1, \ldots, m_i\}$ denote the set of Web pages that point to W_i, $B_i \subseteq \Omega$,

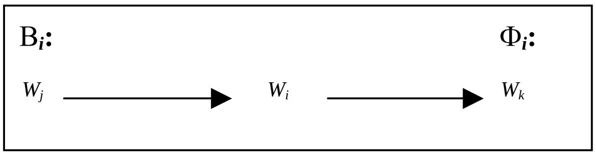


Figure 1. Web pages and links as viewed for PageRank.

It can be seen that this view models the Web as a directed graph denoted by, say, G.

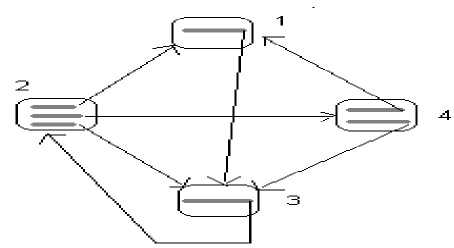
2.2.3. The Fundamental Equation

Based on the extended citation principle and using the graph model, the PageRank value of a Web page W_i, denoted by R_i, is defined using the following fundamental equation [19, 26]:

$$R_i = \sum_{W_j \in B_i} \frac{R_j}{L_j} \qquad (1)$$

where L_j denotes the number of outgoing links from the page W_j. Equation 1 is a homogenous and simultaneous system of linear equations, which, as it is well-known, always has trivial solutions (the null vector), but which has nontrivial solutions too if and only if its determinant is equal to zero. Another commonly used technical definition of PageRank is as follows. Let $G = (V, A)$ denote the directed graph of the Web, where the set $V = \{W_1, W_2, \ldots, W_i, \ldots, W_N\}$ of vertices denotes the set of Web pages. The set A of arcs denotes the links (given by URLs) between pages. Let $M = (m_{ij})_{N\times N}$ denote a square matrix attached to graph G such that $m_{ij} = 1/L_j$ if there is a link from W_j to W_i, and 0 otherwise. Because the elements of matrix M are the coefficients of the right hand side of equation 1, this can be re-written in matrix form as $M \times R = R$, where R denotes the vector of PageRank values, i.e., $R = [R_1, \ldots, R_i, \ldots, R_N]^T$. If the graph G is strongly connected, the column sums of the matrix M are equal to 1 (stochastic matrix). Because the matrix M has only zeroes in the main diagonal, the matrix $M - I$ has zero column sums (I denotes the unity matrix; 1 is subtracted from the main diagonal of the matrix M). Let D denote this determinant, i.e., $D = |M - I|$. If every element of, say, the first line of D is doubled we get a new determinant D' and we have $D' = 2D$. We add now, in D', every other line to the first line. Because the column sums in D are null, it follows that $D' = 2D = D$, from which we have $D = 0$. The matrix $M - I$ is exactly the matrix of equation 1, hence it has nontrivial solutions too (of which there are an infinity). The determinant $|M - I|$ being equal to 0 is equivalent to saying that the number 1 is an eigenvalue of the matrix M.

The PageRank values are computed in practice using some numeric approximation procedure to calculate the eigenvector R corresponding to the eigenvalue 1 or to solve the fundamental equation 1. Computational details as well as convergence considerations in the numeric algorithms used are presented and discussed in [1, 19, 20, 24]. Figure 2 shows an example. The existence of isolated vertices (Web pages without outgoing and incoming links), sink vertices (Web pages with incoming links but without outgoing links), and source vertices (Web pages with outgoing links but without incoming links) cannot be excluded in the real Web.



Matrix M					PageRank
0	1/3	0	1/2		0.325
0	0	1	0		0.651
1	1/3	0	1/2		0.651
0	1/3	0	0		0.217

Figure 2. A small Web, its graph *G*, matrix *M*, and PageRank values (eigenvector corresponding to eigenvalue 1).

Equation 1 has a number of versions which are practical enhancements of it in order to cope with WWW realities (e.g., loops, sink pages, etc.). In [3, 4, 27], several versions to compute PageRank values for all accessible Web pages (whose graph is not necessarily strongly connected) are given. These versions can be written in a compact form as follows:

$$R_i = \alpha \cdot (1 - d) + \beta \cdot d \cdot \sum_{W_j \in B_i} \frac{R_j}{L_j} + \gamma \cdot E \qquad (2)$$

Different versions are obtained for the following values of the parameters α, β and γ as follows (Table 1):

Table 1. Table of parameters whose values determine different versions of PageRank's fundamental equation.

α	β	γ	**Version**
0	$1/d$	0	Fundamental equation 1
0	1	0	version
1	1	0	version
$1/N$	1	0	version
0	d/c	0	version
0	d/c	1	version

where $0 \le d \le 1$ is a constant called damping factor (usually set to 0.85), N denotes the total number of Web pages, and c is a normalisation factor. Note that, quite understandably, few details are given exactly as regards d and c, which are important parameters but do not alter the extended citation principle or the meaning of the fundamental equation 1.

As it was initially suggested [4], the PageRank value of a Web page can be interpreted using an easily bored random surfer metaphor: the probability that a random surfer will be at a certain page after following a large number of forward links. Thus, all the PageRank values form a probability distribution over the Web, so that the probabilities sum up to unity. The solutions of the fundamental equation 1 can be so scaled as to satisfy this condition. Thus, in Figure 2, the values that comply (within an inherent numeric approximation error) with this probabilistic interpretation are as follows: 0.163, 0.326, 0.326, 0.109 (this is an eigenvector, too, and they are roots of equation 1 as well, and proportional to the eigenvector corresponding to eigenvalue 1). A stochastic view of PageRank based on Markov chains is detailed in [2].

The PageRank method has proved to be useful for other purposes, too, for example as an importance metric in crawling the Web pages [7, 8].

3. INTERACTION INFORMATION RETRIEVAL

The Interaction Information Retrieval (I^2R) paradigm was proposed in [11]. I^2R exploits the changing nature of links between objects, and calculates their relative importance as activity levels. [12] contains a more detailed description of the theoretical and practical aspects of I^2R, whereas [13] presents and treats I^2R within a wide formal context of IR models.

3.1. Retrieval and Ranking

The retrieval and ranking of objects follows a usual IR scenario, and is performed in several steps as follows:

(a) Find the object-documents containing terms of the object-query, and link the object-query with these object-documents.
(b) Compute the importance of objects.
(c) Rank the objects according to their importance.

The object-documents form an interconnected network of artificial neurons. The query is integrated into this network as any other object would. The importance of objects is expressed by their activity levels. Retrieval and ranking are based on activation spreading starting from object-query according to a winner-takes-all strategy. The object-documents belonging to reverberative circles are retrieved (local memories evoked by the object-query).

3.2. I^2R Method

The I^2R method uses a connectionist approach based on dynamical systems, and it provides a way to compute the relative importance of objects as activity levels based on a qualitative model from which quantitative (implementable, numeric) models can be obtained.

3.2.1. Qualitative Model of I^2R

The qualitative model of I^2R consists of a principle, model and generic equation (from which specific computational models can be derived).

3.2.1.1. The Principle

The underlying idea of ANNs (Artificial Neural Networks) goes back to [23], where it is stated, as a fundamental law or principle, that the amount of activity of any artificial neuron depends on its weighted input, on the activity levels of artificial neurons connecting to it, and on inhibitory mechanisms. This idea gave birth to a huge literature and many applications, especially due to results obtained in, e.g., [14, 18, 22].

The qualitative model of I^2R is based on the above principle which it applies to IR. Inhibitory mechanisms are not assumed, and the principle of I^2R can be formulated as follows: the activity level of an object is determined by the activity levels of objects which are linked to it.

3.2.1.2. The Model

In order to apply the I^2R method in practice, objects (e.g., documents, Web pages) are assigned (or modelled as) artificial neurons, these form an ANN. Let (Figure 3)

(i) $\Delta = \{O_1, O_2, \ldots, O_i, \ldots, O_N\}$ denote a set of objects (e.g., documents, Web pages); each object O_i is assigned an artificial neuron $\aleph_i$, $i = 1, \ldots, N$; thus we may consider $\Delta = \{\aleph_1, \aleph_2, \ldots, \aleph_i, \ldots, \aleph_N\}$,

(ii) $\Phi_i = \{\aleph_k \mid k = 1, \ldots, n_i\}$ denote the set of artificial neurons that are being influenced (i.e., synapsed) by $\aleph_i$, $\Phi_i \subseteq \Delta$,

(iii) $B_i = \{\aleph_j \mid j = 1, \ldots, m_i\}$ denote the set of artificial neurons that influence (i.e., synapse to) $\aleph_i$, $B_i \subseteq \Delta$.

$$\mathbf{B}_i\textbf{:} \qquad\qquad\qquad\qquad \mathbf{\Phi}_i\textbf{:}$$
$$\aleph_j \longrightarrow \aleph_i \longrightarrow \aleph_k$$

Figure 3. Objects and links as viewed in I^2R.

3.2.1.3 .The Fundamental Equation

The fundamental equation of the I^2R method is the general network equation:

$$\mu_i \frac{dz_i(t)}{dt} = I_i(t) - z_i(t) + \sum_{\aleph_j \in B_i} f_j(z_j(t), w_{ij}, z_i(t)) \quad (3)$$

where

- t denotes time,
- $z_i(t)$ denotes the activity level of the ith artificial neuron,
- w_{ij} denotes the weight of a link from the jth to the ith artificial neuron,
- $I_i(t)$ denotes external input to the ith artificial neuron,
- $f_j(z_j(t), w_{ij}, z_i(t))$ denotes the influence of jth artificial neuron upon the ith artificial neuron,
- μ_i coefficient.

Equation 3 is a simultaneous system of differential equations of the first degree. This is a generic equation, and can have different forms depending on the choice of I_i, f_j, w_{ij} and μ_i corresponding to the particular case or application where the ANN is being used. When applied to neurons μ_i denotes a membrane time constant, z_i denotes membrane voltage, I_i means an external input, w_{ij} is interpreted as a weight associated to the synapse, whereas f_j takes the from of a product between the weight and z_j. For analogue electric circuits, the time constant μ_i is a product between resistance and capacitance, z_i denotes the potential of a capacitor, the left hand side of the equation is interpreted as a current charging a capacitor to potential z_i, whereas the summed terms mean potentials weighted by conductance [10].

3.2.2. A Quantitative Model of I^2R

A quantitative model of I^2R is a computational model derived from the qualitative model, and can be numerically implemented by an algorithm. One possible quantitative model, which has proved useful in several applications, is briefly described below.

Because the objects to be searched are IR objects, e.g, documents, no external input is assumed, so we take $I_i(t) = 0$. One way to define f_j explicitly is to conceive the influences of object j upon object i as being determined by the strengths of the connections which convey this influence, i.e., weights w_{ij} of the links between them, whereas the coefficient μ_i can be taken as being equal to unity. Equation 3 thus reduces to the following equation:

$$\frac{dz_i(t)}{dt} = -z_i(t) + \sum_{\aleph_j \in B_i} w_{ij} \quad (4)$$

The following computation is applied for the weights w_{ij} (but other methods may also be applied). Each $\aleph_i$ is associated an n_i-tuple of weights corresponding to its identifiers (e.g., keywords) t_{ik}, $k = 1, \ldots, n_i$. Given now another $\aleph_j$. If identifier t_{jp}, $p = 1, \ldots, m_j$, occurs f_{ijp} times in O_i then there is a link from $\aleph_i$ to $\aleph_j$, and this has the following weight:

$$w_{ijp} = \frac{f_{ijp}}{n_i} \tag{5}$$

Formula 5 can be applied in a binary (i.e., $f_{ijp} = 1$ or 0) or non-binary form. If identifier t_{ik} occurs f_{ikj} times in O_j, and df_{ik} denotes the number of objects in which t_{ik} occurs, then there is a link from $\aleph_i$ to $\aleph_j$, and this has the following weight (inverse document frequency):

$$w_{ikj} = f_{ikj} \cdot \log \frac{2N}{df_{ik}} \tag{6}$$

The total input to $\aleph_j$ is then given by

$$\sum_{k=1}^{n_i} w_{ikj} + \sum_{p=1}^{n_j} w_{ijp} \tag{7}$$

It can be shown that (i) the solutions of equation 4 are of the form $Ke^{-t}(e^t - 1)$, where K is a constant, and that (ii) when the network operates for retrieval (i.e., activation spreading according to WTA), the activity level of an object is directly proportional to its total input. The quantitative model described above can be deduced from other forms of the fundamental equations, too.

An interesting property of this model is that it is able to return important documents even if they do not contain any of the query terms, but which are strongly linked with documents containing query terms. This property also appears in recent Web searching models [25].

4. PAGERANK: ANOTHER QUANTITATIVE MODEL OF I^{2}R

It will be shown that the PageRank method can be conceived as another quantitative model of the I^{2}R model, i.e., it can be deduced from the qualitative I^{2}R model.

4.1. The Principles

The principles on which PageRank and I^{2}R are based are equivalent with each other due to the following reasons:

(a) In PageRank, the importance of a Web page is expressed by its citation level, whereas in I^{2}R the importance of an object is given by its activity level, and thus PageRank's concepts of citation level and I^{2}R's concept of activity level correspond to each other, they can be paralleled.

(b) In PageRank, the citation level of a Web page depends on the citation levels of the pages pointing to it (and thus implicitly also on their number), whereas in I^{2}R the activity level of an object depends on the activity levels of the objects linking to it (and thus implicitly also on their number). Thus the importance of pages/objects depends on equivalent factors in the two models.

Hence, the principle of PageRank and that of I^{2}R are or can be made equivalent with each other.

4.2. The Models

The models used in both PageRank and I^{2}R are the same, see parts 2.2.2 and 3.2.1.2).

4.3. The Formulas

In I^{2}R, the activity levels are conceived as dynamical quantities which vary with time during operation. In PageRank, the citation levels are static quantities: once computed they remain constant while being used (in the retrieval process). If I^{2}R's activity level is viewed as a particular case, namely constant in time, then I^{2}R's fundamental equation 3 has a null in its left hand side (the derivative of a constant is zero), and hence is asking for finding the equilibrium as a solution. Thus, it becomes:

$$0 = I_i(t) - z_i(t) + \sum_{\aleph_j \in \mathrm{B}_i} f_j(z_j(t), w_{ij}, z_i(t)) \tag{8}$$

No external inputs to Web pages are assumed in PageRank, hence we take $I_i = 0$. It can be seen that, from a numerical point of view, in PageRank, the citation level of a Web page is inversely proportional with the number of links that pages pointing to that page have (if the Web pages have one link each then the matrix M has binary values, the inverse proportionality becomes visible if at least on Web page has more than one link, and L_j appears in denominators). Let the identifier associated to an object be an URL as such. In this case the binary version of formula 5 is identical with the term $1/L_j$ of PageRank's fundamental equation 1: the link from object j to object i has weight $w_{jik} = f_{jik}/n_j$, and $t_{jik} = \mathrm{URL}_i$ (i.e., object j contains the URL address of object i), $f_{jik} = 1$, $n_j = L_j$. Thus, I^{2}R's fundamental equation re-writes as follows:

$$0 = -z_i(t) + \sum_{\aleph_j \in \mathrm{B}_i} f_j(z_j(t), \frac{1}{n_j}, z_i(t)) \tag{9}$$

Taking into account the principle of PageRank (or, equivalently, of I^{2}R), the function f_j does not depend on z_i (the citation level of a Web page does not depend on its own citation level), but it depends on z_j (the citation level of a Web page does depend on the citation levels of the pages linking to it). As it is common to take in ANNs, the function f_j is taken as the dot product of the vector of activity levels and corresponding weights of the objects pointing to it (linear combination), i.e., $f_j = z_j w_{ij}$. The fundamental equation of I^{2}R re-writes now as follows:

$$0 = -z_i(t) + \sum_{\aleph_j \in \mathrm{B}_i} z_j(t) w_{ij} \tag{10}$$

Because z_j does not depend on time (see above), equation 10 becomes:

$$z_i = \sum_{\aleph_j \in \mathrm{B}_j} \frac{z_j}{n_j} \tag{11}$$

which is the same as the fundamental equation 1 of PageRank.

Versions of PageRank's equation can also be obtained. If the influence function f_j of equation 9 is defined as $f_j = d \cdot z_j \cdot w_{ij}$, $0 < d \leq 1$, then the following version of PageRank's equation is obtained:

$$z_i = d \cdot \sum_{O_j \in \mathrm{B}_j} \frac{z_j}{n_j} \tag{12}$$

If the external input I_i of equation 8 is defined as $I_i = 1 - d$, $0 < d < 1$, and the influence function f_j as $f_j = d \cdot z_j \cdot w_{ij}$, the following version of PageRank's equation is obtained:

$$z_i = 1 - d + d \cdot \sum_{O_j \in \mathrm{B}_j} \frac{z_j}{n_j} \tag{13}$$

The PageRank values constitute the equilibrium points of the system of the interlinked Web pages, whose Jacobian matrix is as follows:

$$J = \left(\frac{\partial f}{\partial z_j} \right)_i = M - I \tag{14}$$

Because the number 1 is an eigenvalue of the matrix *M* (see part 2.2.3), the Jacobian matrix *J* is singular. The Hessian of the system is singular, too (*J* has constant elements), hence all the second partial derivatives are zero. Thus, we may say that the PageRank values constitute a neutral equilibrium point of the Web. As it is well-known, from a neutral equilibrium the system may jump to another such equilibrium, i.e, when new PageRank values are being computed the Web is being moved from one neutral equilibrium to another.

5. CONCLUSIONS

After describing the PageRank and I^2R methods (the principles on which they are based, the formal models they use, and the fundamental equations they are built on) it was shown that the PageRank method can be conceived as a particular quantitative model of the I^2R method: their principles and models are the same, and PageRank's fundamental equation can be obtained as a particular case of I^2R's fundamental equation.

This makes it possible to view PageRank from a different perspective (beside the usual stochastic one), namely as a particular connectionist (dynamic) system which looks for equilibrium in the state space.

6. ACKNOWLEDGEMENTS

The author would like to thank Zsolt Tuza for helpful discussions, and to acknowledge the support of grants OTKA T 037821 (National Foundation fro Scientific Research, Hungary), NKFP OM-00359/2001 (Ministry of Education, National Research and Development Program, Hungary), and AKP 2001-140 (Hungarian Academy of Sciences).

7. REFERENCES

[1] Arasu, A. (2002). PageRank Computation and the Structure of the Web: Experiments and Algorithms. *Proceedings of the World Wide Web 2002 Conference*, Honolulu, Hawaii, USA, 7-11 May, http://www2002.org/CDROM/poster (visited: 4 Nov 2002)

[2] Breyer, L.A. (2002). Markovian Page Ranking Distributions: Some Theory and Simulations. http://www.lbreyer.com/preprints/googlerip.pdf (visited: 4 Nov 2002)

[3] Brin, S., Motwani, R., Page, L., and Winograd, T. (1998). What can you do with a Web in your Pocket?. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, http://www.n3labs.com/pdf/brin98what.pdf (visited: 4 Nov 2002)

[4] Brin, S., and Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Proceedings of the 7th World Wide Web Conference*, Brisbane, Australia, 14-18 April, pp: 107-117

[5] Brin, S., and Page, L. http://www.google.com

[6] Carriere, J.,and Kazman, R. (1997). Webquery: Searching and Visualising the Web Through Connectivity. Proceedings of the World Wide Web Conference, Santa Clara, California, April, pp: 701-711

[7] Chakrabarti, S., Dom, B.E., Gibson, D., Kleinberg, J. (1999). Mining the Link Structure of the World Wide Web. *IEEE Computer*, Vol. 32, No. 8, pp: 60-67

[8] Cho, J., Garcia-Molina, H., and Page, L. (). *Efficient Crawling Through URL Ordering*. Stanford University, http://www-db.stanford.edu/~cho/crawler-paper (visited: 4 Nov 2002)

[9] Crestani, F., and Lalmas, M. (2001). *Logic and Uncertainty in Information Retrieval*. European Summer School in Information Retrieval, Milan, Italy. (Also: Agosti, M., Crestani, F., and Pasi, G. (eds.) *Lectures on Information Retrieval*, Lecture Notes in Computer Science, LNCS 1980, Springer Verlag, pp: 179-206)

[10] De Wilde, Ph. (1996). *Neural Network Models*. Springer.

[11] Dominich, S. (1994). Interaction Information Retrieval. *Journal of Documentation*, vol. 50, no. 3, pp: 197-212

[12] Dominich, S. (1997). The Interaction-based Information Retrieval Paradigm. In: Kent, A., and Williams, G.W. (eds.) *Encyclopedia of Computer Science and Technology*. Marcel Dekker, Inc., New York – Basel – Hong Kong, vol. 37, suppl. 22, pp: 175-193

[13] Dominich, S. (2001). *Mathematical Foundations of Information Retrieval*. Kluwer Academic Publishers, Dordrecht, Boston, London.

[14] Feldman, J.A., and Ballard, D.H. (1982). Connectionist models and their properties. *Cognitive Science*, 6, pp: 205-254

[15] Garfield, E. (1955). Citation indexes for science. *Science*, p. 108

[16] Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation. *Science*, pp: 471-479

[17] Google, Inc., http://www.google.com

[18] Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybertnetics*, 23, pp: 121-134

[19] Haveliwala, T.H. (1999). *Efficient Computation of PageRank*. Stanford University, http://dbpubs.stanford.edu:8090/pub/1998-31 (visited: 4 Nov 2002)

[20] Haveliwala, T.H. (2002). Topic-Sensitive PageRank. *Proceedings of the World Wide Web Conference 2002*. May 7-11, Honolulu, Hawaii, USA, http://www2002.org/CDROM/ (visited: 4 Nov 2002)

[21] Henzinger, M.R., Heyden, A., Mitzenmacher, M., and Najork, M. (1999). Measuring Index Quality Using Random Walks on the Web. *Proceedings of the 8th International World Wide Web Conference*, Toronto, Canada, pp: 213-225

[22] Hopfield, J.J. (1984). Neurons with graded response have collective computational properties like those of two-states neurons. *Proceedings of the National Academy of Sciences*, 81, pp: 3088-3092

[23] James, W. (1890). *Psychology (Briefer Course)*. New York: Holt, Chapter XVI, “Association”, pp: 253-279

[24] Kim, S.J., and Lee, S.H. (2002). An Improved Computation of the PageRank Algorithm. In: Crestani, F., Girolamo, M., and van Rijsbergen, C.J. (eds.) *Proceedings of the European Colloquium on Information Retrieval*. Springer LNCS 2291, pp: 73-85

[25] Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. http://www.cs.cornell.edu/home/kleinber/auth.pdf (visited: 4 Nov 2002)

[26] Page, L. (2001). United States Patent 6,285,999. September 4 (visited: 4 Nov 2002)

[27] Page, L., Brin, S., Motwani, R., and Winograd, T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web*. Stanford University, http://dbpubs.stanford.edu:8090/pub/1998-66 (visited: 4 Nov 2002)