Towards of the Construction of a Global Bibliometric Indicator

Marios Poulos*, Nikolaos Korfiatis** and George Bokos*

*mpoulos@ionio.gr, gbokos@ionio.gr,
Dept. of Archives and Library Sciences Ionian University, Palea Anaktora 49100, Corfu (Greece)

** nk.inf@cbs.dk <nk.inf@cbs.dk>
Dept of Informatics, Copenhagen Business School. Copenhagen (Denmark)

Introduction

It's conventional wisdom that the availability of scholar related information available on the World Wide Web has been dramatically increased. At the same time the Web becomes the main medium of publication and distribution of information. The results of surveys lead us to the understanding, improvement and the appropriate exploitation of the Web for the scientific research. Methods, techniques and tools originated in the application of bibliometric studies in traditional scientific publications, have affected the majority of bibliometric methods (Bar-Ilan et al. 2002).

The most significant bibliometric techniques which are used in the above applications are: Citation Analysis, Scientometrics, Scholarly Communication Studies, Infometrics, Web-metrics, Social Network Analysis (Bar-Ilan et al. 2002, Bjorneborn et al. 2001, Casserly et al. 2003). However, as Snizek indicates, the problems of the above indicators are focused on a multi-parametric variable number of interdependent values (Snizek, 1995). In this paper we introduce a method in which we evaluate each paper with regard to two parameters. First, with respect to the journal in which this paper has been published and second, with respect to the value of the cited references of this paper. These indicators are presented as follows. For the evaluation of the value of a published journal the impact factor of this journal is used with often some negative examples (Smith, 1997). For the evaluation of each of the cited references of the paper in question we take into account the following indicators: (a) the impact factor of the published journal. (b) The degree of the significant relation (SR) of the scientific area, which is calculated between the cited and the published journals (Gomez et al. 1999). (c) The degree of the significance of the social relation between the authors of the published and the cited references as indicated by metadata schemas (Brickley et al. 2006, Korfiatis et al. 2006).

To this end in this research in progress paper we attempt to take into consideration the above factors in order to conceptualize a theoretic evaluation model. We used a technique of the fuzzy set theory in which all the above best indicators should be added in a general mathematic matrix. Finally, the global indicator of each paper is calculated with regard to the best integral model using the cross-correlation test.

Method

The proposed global matrix creates a sophisticated schema, where each indicator is graduated by five fuzzy logic concepts. In particular, we consider a membership relation $\mu_A(u)=1$ in which "u" is an element of the set A, and $\mu_A(u)=0$ in the case "u" is not an element of the set A. The degree of membership is defined by a through a generalized characteristic function called membership function $\mu_A(u):U\rightarrow[0,1]$, where U is called the universe, and A is a fuzzy subset of U. The values of the membership function are real numbers in the interval [0,1], where 0 means that the category of the bibliometric indexes is not a member of the set and 1 means that it belong entirely (N. Kasabov 1996). One way of defining a membership functions is though an analog function.

The representing value for a fuzzy set transformation is given with a Boolean type value yes (1) or no (0). Thus a combined number of values [0,0,0,0,0] may correspond to the value 0, while a combined number of values [1,1,1,1,1] corresponds to the value 1. The other combination, for example [0,1,1,1,1] or [1,1,0,1,1], etc, gives value 0.75. At this stage we constructed an ideal bibliometric indicators set [1,1,1,1,0] which represents the max fuzzy set value for a citation.

The Implementation of the Method

This section is divided in four stages.

In the first stage, we create a theoretic article, which is represented using a matrix (see equation 1), which has 5xM dimensions, where the M is the maximum number of the published articles which belong to the same thematic area with the investigated paper.

$$A=\begin{matrix}\mu_{A_{11}}(u) & \cdots & \mu_{A_{M1}}(u)\\ \mu_{A_{12}}(u) & \cdots & \mu_{A_{M2}}(u)\\ \mu_{A_{13}}(u) & \cdots & \mu_{A_{M3}}(u)\\ \mu_{A_{14}}(u) & \cdots & \mu_{A_{M4}}(u)\\ \mu_{A_{15}}(u) & \cdots & \mu_{A_{M5}}(u)\end{matrix} \quad (1)$$

Thus, a vector *C=[A]* of *1x5xM* dimension is constructed, which is then used in the next stage.

Thus, each column of the proposed matrix represents the publication, which cites the investigated article. In total, the above matrix depicts a degree of representation of all the citations of the investigated paper with regard to the total number (M) of the publications in the same subject area. In any case, the number of columns is completed. In the empty registries we consider that these are represented by the vector (0,0,0,0,0).

In the second Stage, In this stage, we conceptualize an ideal theoretic article, which is represented using a matrix A_1 (see equation 1).

$$A_1 = \begin{matrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{matrix} \quad (2)$$

Then the coordinate vector of the above matrix is:

Thus, a vector $\hat{C}_x = [A_1]$ of 1x5M dimension is constructed, which is then used in the next stage.

For example we adopted the number M=10^4. Thus, the vector $\hat{C}_x$ has **(**1x10^4) dimensionality.

In the third stage, we constructed a matrix, labeled A_2 (see equation. 5) which represents an article, which has been cited four times. The citations of this article are represented by the first (4) four columns and the other columns are completed by (M-4) or (10^4-4) vectors which contain zero elements (0, 0, 0, 0, 0).

$$A_2 = \begin{matrix} 0.25 & 0 & 0.75 & 1 & 0 \\ 0.50 & 0.75 & 0.50 & 0.25 & 0 \\ 0 & 0 & 0.25 & 0.50 & 0 \\ 0.75 & 0.25 & 0.50 & 1 & 0 \\ 0 & 0.50 & 1 & 0.25 & 0 \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} . \\ . \\ . \\ . \\ . \end{matrix} \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix} \quad (3)$$

Then, the coordinate vector of the above matrix is:

Thus, a vector $\hat{C}_x = [A_2]$ of *1x5xM* or for example 1x10^4 dimensionality is constructed.

In fourth stage the extracted set of auto-correlation coefficients $\hat{C}_x$ of an ideal-cited paper case were submitted to the cross-correlation procedure (Morisson et al 1976) along with another cited paper $\hat{C}_y$ as described below:

$$r = \frac{\sum_{i=1}^{5*10^4-1} (\hat{C}x_i - \bar{\hat{C}}x_i)(\hat{C}y_i - \bar{\hat{C}}y_i)}{\sqrt{\sum_{i=1}^{5*10^4-1} (\hat{C}x_i - \bar{\hat{C}}x_i)^2 (\hat{C}y_i - \bar{\hat{C}}y_i)^2}} \quad (4)$$

The extracted cross-correlation coefficient is a number between -1 and 1, which measures the degree to which two variable sets are linearly related.

Conclusions and Further Research

In this paper, we attempted to produce a global bibliometric index for a published article. Thus, we constructed a global ideal article with (5) five high score indicators for each citation. For this implementation, we used the fuzzy logic theory in order to classify in (5) five degrees the bibliometric indicators. Finally, we extracted a bibliometric indicator for a simple article by correlating this with the ideal article. This paper may be considered as the first attempt to construct the proposed IGBI indicator and thus the research presented here should be considered as research in progress. Our major concern is to evaluate the IGBI index against well known comparisons such as journal status (Bollen et al. 2006) using an extensive citation graph. Furthermore, the extension of a semantic vocabulary such as COAP and the construction of a citation index to provide real time IGBI analysis is also an important step.

References

Bar-Ilan, J. & Peritz, B. (2002). Informetric theories and methods for exploring the internet: an analytical survey of recent research literature. *Library Trends*, 50 (3), 371-392.

Bjorneborn, L. & Ingwersen, P. (2001). Perspective of webometrics. *Scientometrics*, 50 (1), 65-82.

Brickley, D. and L. Miller.(2006). FOAF Vocabulary Specification" Retrieved July, 2006 (http://xmlns.com/foaf/0.1/).

Bollen, J., Rodriguez, M.A., Van de Sompel, H., (2006). Journal Status, *Scientometrics*, 69, (3)

Casserly, M. & Bird, J. (2003). Web citation availability: analysis and implications for scholarship. *College & Research Libraries*, 64 (4), 300-317.

Gomez, I. &, Maria Teresa Fernandez and Jesus Sebastian (1999). Analysis of the structure of international scientific cooperation networks through bibliometric. *Scientometrics*, 44(3) 441-457.

Korfiatis, N., Poulos, M., & Bokos, G. (2006). Social Metadata for the Impact Factor. *The Electronic Library (TEL) Journal.* (Forthcoming)

Morrison, N., Donald, F., 1976. Multivariate Statistical Methods, (New York; McGraw-Hill Book Company,)

N. Kasabov (1996). Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering, MIT Press,. Fuzzy systems. (sections 3.1.1-3.1.2, pp.167-169).

Smith R. (1997). Journal accused of manipulating impact Factor, *British Medical Journal*, 314, N. 7079 , 463-466.

Snizek W. E. (1995). Some observations on the use of bibliometric indicators in the assignment of university chairs. *Scientometrics* 32(2), 177-120.