A New Framework for the Citation Indexing Paradigm*[1]

Dimitris A. Dervos[(1)], Nikolaos Samaras[(2)], Georgios Evangelidis[(2)], and Theodore Folias[(2)]

[(1)]Correspondence author: Information Technology Dept., Alexander Technology Educational Institute (ATEI) of Thessaloniki, P.O. BOX 141, 57400 Sindos, Greece
[(2)]Dept. of Applied Informatics, University of Macedonia, P.O. BOX 1591, 54006 Thessaloniki, Greece
Email: [(1)]dad@it.teithe.gr, [(2)]{samaras, gevan, folias}@uom.gr

ABSTRACT

A new citation indexing paradigm is proposed: the cascading citation indexing framework (c^2IF, for short). It improves the way research publications are assessed for their impact in promoting science and technology. Given a collection of articles and their citation graph, citations are considered at the (*article*, *author*) level. Each one article is uniquely identified by means of the Digital Object Identifier (DOI, http://www.doi.org). To identify each one author uniquely, a Universal Author Identifier (UAI) scheme is established. In addition to the citations directly made to a given (article, author) pair, citation paths that target each one citing article are also considered. The granularity of the paradigm is further increased by introducing the concept of the *chord*, whereby a citation path of length one co-exists with paths of length two or higher, involving the same source- and target- articles. The c^2IF output emerges in the form of a medal standings table, analogous to the one that ranks teams at athletic events: when two (article, author) pairs receive the same number of (direct) citations, the one that is cited by more popular articles (i.e. articles that comprise targets to a larger number of paths in the citation graph), is assigned a higher rank value.

INTRODUCTION

Nowadays, developments like the evolving scholarly communication environment, the open access movement, and the globalization in academia and research advance with a rapid pace. As a result, more intense becomes the need for an improved scheme that assesses the contribution research publications, authors, and scientific collections make in promoting science and technology. Eugene Garfield (1955, 1999, and 2005) has proposed the use of the journal impact factor metric. The impact factor (codenamed ISI IF in the following) is a standardized metric that can be used to measure the way a journal/conference receives citations on its articles over time. Two more metrics of this type are the immediacy index (Tomer, 1986) and the cited half-life (Glänzel, and Moed, 2002). The ISI IF approach ranks article in accordance with the prestige value of the conference or journal where it is published in. Although such a metric comprises a useful indicator of scholarly status, concerns have been expressed over the usefulness and the fairness of its implementation (Coleman, 2006; Moed, 2005; Hoeffel, 1998; Smith, 1981). Bollen, Rodriguez and Van de Sompel (2006) in particular, note that ISI IF focuses on the popularity of the cited item, ignoring the prestige value of the citing one. In this respect, it is rendered to be impossible to apply in many other areas, such as for example the WWW. At this point, it is worth noting that when it comes to popularity and prestige value assessment, the web page paradigm has a lot in common with the research article one, when the latter is considered in the context of the open access movement, today.

* *Research conducted along the lines of the Cascading Citation Analysis Project (C-CAP, http://www.ccapnet.org), funded by the Research Committees of ATEI, and the University of Macedonia, Thessaloniki, Greece*

When citations are considered at the published article level, the article's scholarly value is measured by utilizing two major metrics: (a) the number of direct citations received, and (b) the impact factor of the hosting conference/journal. The first metric reflects the popularity of the particular article, since a large number of citations received usually implies significant contribution in the corresponding scientific field. The second metric quantifies the scholarly credibility of the article in question, since acceptance by a widely recognized conference or journal most probably signifies the presence of a pioneering character and expert recognition in what is being reported. Consequently, articles published in high impact factor journals/conferences reach a broader audience, and they are likely to receive a larger number of citations.

Today, thanks to the open access movement, worldwide accessibility of research articles does not necessarily mean that the latter are published with prestigious conferences or journals. In addition, the ISI IF metric is reported to vary considerably from one scientific field to another (Moed et al., 1985) The need for an improved methodology that measures the popularity and the scholarly credibility of the published works is justified by the fact that today educational and research institutes utilize such metrics when deciding for compensation levels of researchers or research funding (Kleijnen J.P.C. and Van Groenendaal, 2000).

As an alternative to the ISI IF approach, Bollen, Rodriguez and Van de Sompel (2006) apply a weighted variation of Google's PageRank algorithm (Pinski and Narin, 1976; Brin and Page, 1998). The latter is not meant to replace the former since it assesses the scholarly status on the basis of prestige, as opposed to popularity which is the case for the ISI IF approach. The weighted PageRank algorithm is recursive in its nature; in addition, we note that it can be applied not just for measuring the scholarly status of journals (as it is done in Bollen, Rodriguez and Van de Sompel, 2006) but also that of published articles.

Our approach is analogous to the one of the weighted PageRank algorithm in that citation paths of length greater than one are being exploited. In this respect, the scholarly status is assessed not just in terms of popularity of the cited item (expressed by the number of direct citations received), but also in terms of the prestige of the citing item(s) (expressed by the number of indirect citations received). One first thing to note is the harmonic co-existence of the popularity and prestige measuring metrics in a single output: popularity directly relates to the number of citation paths of length one, having the cited item as their target, and prestige relates to citation paths of length two or larger, having the cited item as their target. A second thing to note is that citations are considered at the (article, author) level; this is done in order to guarantee fairness in scholarly credit assignment to authors; for example, when article '1' which is co-authored (say) by 'A' and 'B' is cited by article '2', co-authored by 'B', 'C', and 'D', the scheme is taken to represent a valid citation for (1,A), and at the same time comprise a self-citation for (1,B). Last but not least, the aim in our cascading citation indexing framework (codenamed: c^{2}IF) is to enrich the citation indexing paradigm, rather than calculate a single value for some metric (as it is done by the weighted PageRank algorithm). In this respect, the c^{2}IF algorithm calculates an output that is intentionally left un-modulated/un-weighted. The approach is justified by considering today's state of the art in database technology that makes possible the processing of extensive citation data corpora, in search for useful associations, patterns, and rules (Dunham, 2003). More specifically, the c^{2}IF output comes in the form of a medal standings table, analogous to the one that ranks teams at athletic events: when two (*article*, *author*) pairs receive the same number of (direct) citations, the one that is cited by more popular articles (i.e. articles that comprise targets to a larger number of citation paths) is assigned a higher rank value.

The c^2IF paradigm is further enriched by counting the number of *chords* associated to each one citation path of length two or larger. By definition, a chord is an instance where two citation paths (one of length one, and the other of length two or greater) are found to involve the same target, i.e. cited (article, author) pair, at the one end, and the same source/citing article at the other end. The scheme facilitates the assignment of additional scholarly credit to the cited (target) item, since the citing (source) item not only cites the former indirectly (via the longer in its length citation path), but also directly.

The paper consists of four core sections. In 'Universal Author Identifier', the need and the specifications for a web-based environment that will assign unique IDs to authors are considered. Next comes the 'Cascading Citations' section which introduces the basic concepts involved in the proposed cascading citation indexing framework, and comments on some preliminary results obtained. In the 'Design and Implementation Issues' section, the technology used for the development of the pilot implementation is discussed. In 'Expected Impact and Future Work', we address issues relating to the impact the cascading citations indexing framework is expected to have in the shaping of a new everyday professional practice for all the actors involved (authors, libraries, publishers, and public users), as well as the future stages of our research project.

UNIVERSAL AUTHOR IDENTIFIER

Given the fact that the proposed c^2IF scheme considers citations at the (article, author) level, each one article and each one author need be identified uniquely. Today, each one article is uniquely identified by means of its digital object identifier DOI value. There is a clear need for an analogous identification scheme that will apply to authors. The day-to-day operation and maintenance of a unique author identifier (UAI) system comprises a task that can only be undertaken by a publicly accredited organization. The proposed system need be a web-based service where authors will be able to log in and acquire an ID that will remain invariant for life, and it will continuously alleviate discrepancies originating from misspelled names, homonyms, aliases, or name variations (eg. Eugene Garfield, E. Garfield, Eugène Garfield, etc.). Equally important is also the requirement for the UAI system to not be tied to any one single source of interest (publisher, citation database vendor, etc.), and as such have the potential to co-function with a number of citation database systems, the open access eprint archives and repositories included (Hitchcock, 2003).

For the UAI system to succeed and enjoy worldwide applicability, the actors involved need be identified from the start, and have a clear benefit from its use, once the full-scale application becomes operational. In this respect, UAI (when coupled to c^2IF) is seen to mainly involve four actors: (a) the author, (b) the publisher, (c) the library, and (d) the general (public) user. Table 1 summarizes on the UAI actors, their roles and the expected benefit in each one case.

The pilot UAI system implementation is a Java based web application allowing each one author to register/update his/her own metadata content and request a unique identifier that s/he is going to retain and make use of for life. Apart from obtaining his/her unique UAI code, the author specifies the subset of his/her personal (meta)data that become globally available to all interested parties. The system supports the industrial standard interface for other applications to connect to and co-function with, over the Internet. The basic functionality to be supported during the pilot implementation phase allows each one author to: (a) register and obtain his/her personal UAI code, (b) determine the own (meta)data that become publicly available, (c) maintain/update his/her UAI entry data content, (d) issue queries to c^2IF, retrieving information relating to the citations (direct and indirect) and the chords that target (article, author) pairs with 'author' being the individual in question. At a later stage, it will become possible for each one author to make use of the UAI system in order to identify published works that s/he has (co-)authored under a different variation of his/her own name, and request credit on their authorship.

Table 1. UAI Actors and Expected Benefit

Author	Publisher	Library	Public User
Receive credit for his/her published articles, despite the variations of own name, and possible existence of homonyms	Utilize the c^2IF output to identify leading researchers in various discipline areas; identify potential guest editors, reviewers, etc	Conduct citation data analysis (say) at a national level. Identify own nationals who publish in certain discipline areas, etc	Trace a given author's previous works, despite the name variation(s) used
Receive credit for indirect citations and chords to own work	Access the latest, up-to-date contact information of each one author	Improved services to users when conducting author searches	Access the latest, up-to-date contact information of each one author
Own name and contact information become globally available	Make use of the UAI database, as a universal directory of authors, categorized by the discipline area(s) they publish in	Improved services to authors via privileged access to the UAI system	Make use of the UAI database, as a universal directory of authors, categorized by the discipline area(s) they publish in

Figure 1 outlines the combined UAI-c^2IF environment, as it is currently being developed, along the lines of the C-CAP project.

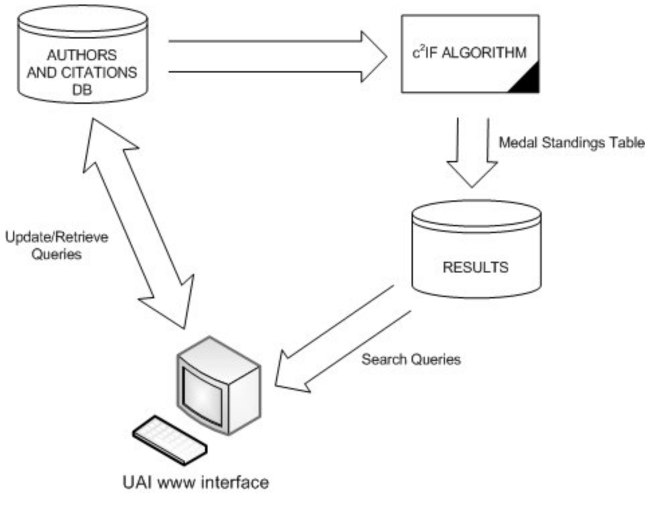

Figure 1. UAI-c^2IF system architecture

CASCADING CITATIONS

Let us consider a small hypothetical collection of five articles labeled, for simplicity, with the integers 1, 2, 3, 4, and 5. Furthermore, let (A, B) be the two authors who have co-authored article 1, A be the author of 2, (B,C) the authors of 3, D the author of 4, and (B,E,F) the authors of 5. A citation graph is a directed graph that represents relationships between articles in terms of citation references. In Figure 2 the citation graph for the hypothetical collection considered is presented. Each one node corresponds to one article. The letters in the box(es) around each node represent the author(s) of the article. References from one article to another are represented by directed arcs. Citations are taken to target (article, author) pairs. For example, (1,A) is cited by 3, along the 3 1 citation path, with 3 being the source and (1,A) being the target of the citation. The latter is said to comprise a 1-gen (direct) citation. In the same manner, 2-gen, 3-gen, …, k-gen citations are defined to be those that target a given (article, author) pair indirectly. For example, (1,A) is cited by 4 via a 2-gen citation, along the 4 2 1 citation path.

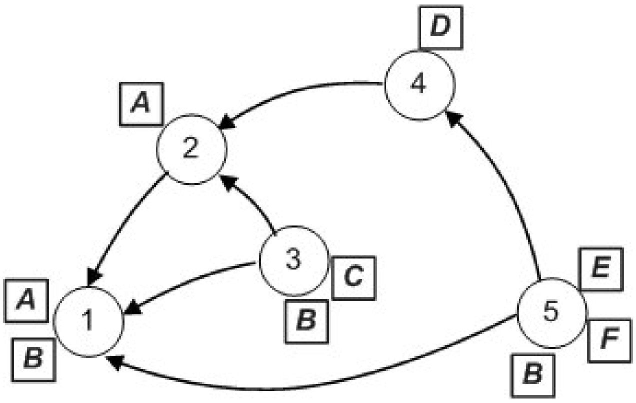

Figure 2. Citation graph of the hypothetical collection

Table 2 lists all the citations present in the citation graph of the hypothetical articles collection considered.

Table 2. Citations, paths, and types present in the hypothetical collection

(author, article)	citation path	citation type
(1,A)	2 1	1-gen
(1,A)	3 1	1-gen
(1,A)	4 2 1	2-gen
(1,A)	5 4 2 1	3-gen
(1,B)	2 1	1-gen
(1,B)	3 1	1-gen
(1,B)	4 2 1	2-gen
(1,B)	5 4 2 1	3-gen
(2,A)	3 2	1-gen
(2,A)	4 2	1-gen
(4,D)	5 4	1-gen

For each one article N, the list of its co-authors is denoted by AL_N. For example, in the hypothetical collection considered: AL_5={B,E,F}. Table 3 summarizes on the symbolism used throughout this paper.

Table 3. Symbolism used

Symbol	Meaning
(N,A)	(article, author) pair (N=1,2,…)
N[A,B]	Article 'N' is co-authored by authors 'A', and 'B'
AL_N	A given article's authors list. Thus, for N[*A*,*B*]: AL_N = {*A*,*B*}

S	The source article of a given k-gen citation path (k=1,2,…)
T	The target article of a given k-gen citation path (k=1,2,…)
S … T	k-gen citation path: S cites T (k=1,2,…)

Self-Citations

Considering the proposed new framework, for the citation indexing paradigm to be complete, useful, and applicable in bibliometrics, self-citations need be identified. Today's practice is to consider citations at the (cited) article level. In this respect, a self citation is said to occur when the set of co-authors of the cited and citing papers are not disjoint (Snyder and Bonzi, 1998). In the new indexing paradigm proposed, authors and articles are each uniquely identified and citations are considered at the (article, author) level. In this respect, a more refined definition of the concept of self-citation now becomes possible to formulate:

Definition: A k-gen (k=1,2,…) citation path S … T represents a self-citation for a given (T,A) pair, when 'A' appears in the authors lists of both the target- and source- articles of the citation path considered (i.e. when $A \in AL_T \cap AL_S$).

For example, considering the citation graph of the hypothetical collection shown in Figure 2, 2 1 represents a 1-gen self-citation on (1,A), 3 2 1 represents a 2-gen self-citation on (1,B), and 5 4 2 1 represents a 3-gen self-citation on (1,B). Also, 3 1 and 5 1 represent 1-gen self citations on (1,B). Apparently, the same citation path may represent (self-)citations to more than one (article, author) pair, without any restriction. Thus, 5 4 2 1 represents a 3-gen self-citation on (1,B) and a 3-gen citation on (1,A).

Glanzel, Thijs, and Schlemmer (2004) suggest that there is no need to exclude self-citations in evaluating bibliometrics. In this respect, self-citations are included in the c2IF output ('Medal Standings Type Output', below).

Chords

For the purpose of increasing the granularity (equivalently: the information content) of the citation indexing paradigm, the concept of the chord is introduced and it is defined as follows:

Definition: A k-gen (k=2,…) citation path S … T represents a chord for a given (T,A) pair, when: (a) $A \in AL_T$, and (b) the path co-exists with a 1-gen citation path involving the same source (S) and target (T) articles.
As in the case of a self-citation, when 'A' appears in the author lists of both the target- and source- articles of the citation path considered, the latter is said to represent a self-chord. For example, 5 4 2 1 in Figure 2 represents a 3-gen chord on (1,A), and a 3-gen self-chord on (1,B).

A chord is considered to be important and worth its inclusion in the citation indexing paradigm for the following reason: the scheme is indicative of an increased probability the target (article, author) pair in question stands in being one of increased impact in promoting science and technology. This is justified by the fact that the source article in question cites the (article, author) target both indirectly (via the k-gen citation), and directly (via the 1-gen citation).

Medal Standings Type Output

Considering the above, the medal standings type tabular output of the c^{2}IF algorithm in the proposed cascading citation indexing framework need be one whereby each one row lists the following: (a) the (article, author) pair in question, (b) the number of 1-gen, 2-gen, …, k-gen citations received, (c) the number of 1-gen, 2-gen, …, k-gen self-citations received (s-citations), (d) the number of 2-gen, …, k-gen chords received, and (e) the number of 2-gen, …, k-gen self-chords received (s-chords). In this respect, the c^{2}IF output for the hypothetical articles collection shown in Figure 2 is presented in Table 4.

Table 4. Medal Standings Output (for the hypothetical collection)

	citations			**s-citations**			**chords**		**s-chords**	
(article, author)	**1-gen**	**2-gen**	**3-gen**	**1-gen**	**2-gen**	**3-gen**	**2-gen**	**3-gen**	**2-gen**	**3-gen**
(1,A)	2	2	1	1	0	0	1	1	0	0
(2,A)	2	1	0	0	0	0	0	0	0	0
(1,B)	1	1	0	2	1	1	0	0	1	1
(4,D)	1	0	0	0	0	0	0	0	0	0
(3,B)	0	0	0	0	0	0	0	0	0	0
(3,C)	0	0	0	0	0	0	0	0	0	0
(5,B)	0	0	0	0	0	0	0	0	0	0
(5,E)	0	0	0	0	0	0	0	0	0	0
(5,F)	0	0	0	0	0	0	0	0	0	0

Preliminary Results

One would ideally expect that a citation graph does not involve any cycles, since each one citing article is expected to be posterior to the one(s) it cites. Yet, this is not always the case; for example, it is possible for a journal preprint to receive a citation from an article that is published at an earlier date than the cited article. Also, it is quite possible to have two articles reference one another. The problem is dealt with by having the c^{2}IF algorithm consider cascading citations recursively up to a pre-specified depth along each one path in the citation graph.

In (Dervos and Kalkanis, 2005), an earlier implementation of the c2IF algorithm was tested against a collection of 1,065,035 citation entries of the CiteSeer database (Giles C.L., Bollacker K., and Lawrence S., 1998). The algorithm ran recursively and considered the cascading citation instances up to k=3 (i.e. 1-gen, 2-gen, and 3-gen), without identifying s-citation, and (s-)chord instances. The results obtained are shown graphically in Figure 3.

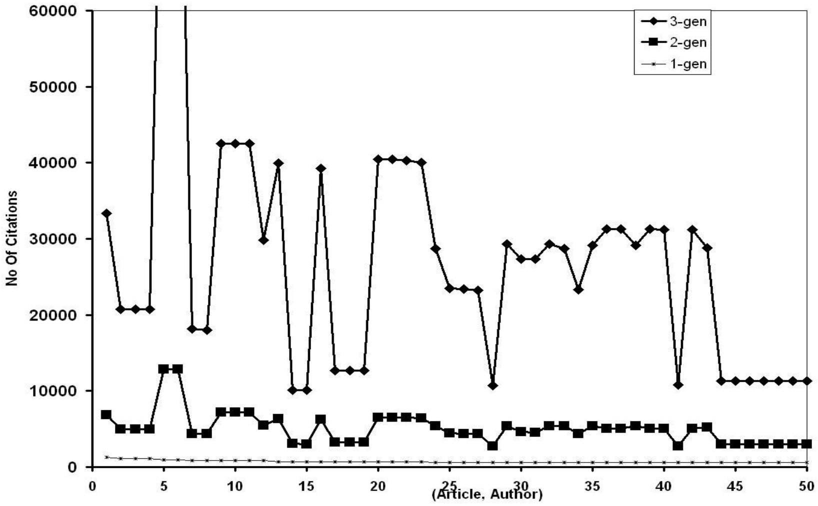

Figure 3. Top 50 (article, author) pairs in the CiteSeer database

The horizontal axis of the graph in Figure 3 registers the rank value assigned to each one (article, author) pair, in accordance with the number of 1-gen citations received. Thus, '1' stands for the (article, author) pair that has received the largest number of 1-gen citations, '2' stands for the second best (article, author) pair, etc.

Commenting on the results shown in Figure 3: the (article, author) entry ranked fifth in accordance with the number of 1-gen citations received, is seen to have attracted almost twice as many 2-gen citations, and almost three times as many 3-gen citations, when compared to the entry ranked first. One additional issue in favor of the former (article, author) entry could be (for example) that the corresponding article was found to have been published twelve years later than the article of the (article, author) entry ranked first. One would normally expect the number of citations received by an (article, author) pair to increase over time.

DESIGN AND IMPLEMENTATION ISSUES

As it has been mentioned already, the pilot UAI system implementation (codenamed: UAI_Sys) is a Java based web application that runs on top of an application server. The Java platform has been chosen for system implementation since it comprises the de facto world-wide standard for developing open source web-based applications, utilizing a large number of available tools and technologies. UAI_Sys is a Java2 Enterprise Edition (J2EE) application that utilizes open source Java tools and technologies provided by the JBoss community (http://www.jboss.org).

The c^2IF algorithm comprises the heart of the proposed system. It utilizes a relational database management system (RDBMS) both for the representation of the citation graph and for the storage of the results (citation paths). For a given positive integer value k, the algorithm computes for each one (DOI, UAI) pair all the 1-gen, 2-gen, …, k-gen (self-)citations, and all of the 2-gen, …, k-gen (self-)chords. Given the above information, the algorithm can easily produce the corresponding medal standings table row entry for the given (DOI, UAI) pair.

Considering the above, the c^{2}IF algorithm has a lot to benefit from an implementation that exploits the inherent parallelism. In this respect, the algorithm is being implemented using the Message Passing Interface (MPI) C++ programming environment.

EXPECTED IMPACT AND FUTURE WORK

The UAI-c^{2}IF system is expected to represent a considerable change to the everyday working life of the four types of actors involved (authors, publishers, libraries, and public users). The UAI component of the web based environment will in effect comprise a 'who-is-who' worldwide database of authors, the content of which is to be maintained up-to-date by the authors themselves.

With guaranteed uniqueness in the identification of each one (article, author) entity in the authors and citations database, the c^{2}IF algorithm will periodically process the citation data, producing an up-to-date version of the medal standings type tabular output. The latter reveals the number of (self-)citations and (self-)chords that target each one (article, author) pair, up to a pre-specified depth in the corresponding paths of the citation graph. This way, authors around the world will be able to monitor not only the direct, but also the indirect (self-)citations, and (self-)chords received by each one article they have (co-)authored. Such information will of course evolve in time, as new citation data are appended to the database.

A set of preliminary c^{2}IF results obtained is indicative of the usefulness of the information obtained by implementing the new cascading citations indexing framework. The fully blown version of the c^{2}IF algorithm is currently being implemented, one that computes all of the (self-)citations and (self-chords) received by each one (article, author) pair, up to the pre-specified depth k. One possible future improvement is the design and development of a weighted variation of the c^{2}IF algorithm, analogous to the one of the Bollen, Rodriguez and Van de Sompel (2006) approach. The scheme is expected to make possible the calculation of a single value reflecting the impact/contribution each one actor represents in the context of the citation data 'space': an actor being an (article, author) pair, an individual article, an author, or a hosting journal/conference. In addition to the calculation of a single impact factor metric, the granularity of the cascading citation indexing paradigm data content facilitates effective analytical processing of the data mining type to be conducted, in order to identify regions of increased research activity, as well as interesting trends in the citations data 'space'.

CONCLUSION

Today, the two main research needs that characterize the majority of the researchers and librarians who use a typical citation indexing environment are the following (Weertman, 2006):

1. To assess a research area to see if it is an active field worthy of entering or pursuing.
2. To evaluate an individual author to help decide whether they would be suitable and relevant to work with, employ, grant funds to, review a manuscript or whether they might be a "rising star" to keep an eye on.

The above, when coupled with today's evolving scholarly communication environment, the open access movement, and the need for effective analytical processing of the citation data in order to identify regions of increased research activity and interesting trends, clearly call for an improved citation indexing paradigm.

The cascading citation indexing framework (c^2IF) increases the granularity of the citation indexing paradigm by:

1. Considering the citations at the (article, author) level.
2. Registering not only the (self-)citations made directly to a given (article, author) target, but also those made indirectly, up to a pre-specified depth along each one path in the citation graph.
3. Introducing the concept of the chord, and opting for the registration of all (self-)chords, up to a pre-specified depth along each one path in the citation graph.

The pilot implementation of the new citation indexing framework includes a universal author identifier subsystem (UAI_Sys) that enables each one author to acquire a unique identification code which remains invariant for life. UAI_Sys is coupled with c^2IF_Sys, the subsystem that processes the citation data, producing a medal standings type tabular output, where next to each one (article, author) entry are listed the numbers of (self-)citations, and (self-)chords received. The c^2IF_Sys output is intentionally left un-modulated/un-weighted in order to: (a) make visible the usefulness of the information revealed by the indirect (self-)citations and (self-)chords received by a given (article, author) target, and (b) facilitate subsequent citation data processing of the data mining type. The aim is of course to have the new framework of the citation indexing paradigm better serve the research needs of the four main actors involved, namely: the author, the librarian, the publisher, and the general (public) user.

ACKNOWLEDGMENTS

The authors wish to thank: (a) Jaakko P. Hyvärinen, and Ypatios Asmanidis, C-CAP programmers, for writing/maintaining the code of the pilot UAI system implementation, (b) Richard Hartley, Anita Coleman, and Peiling Wang, members of the Cascading Citation Analysis Project (C-CAP) Extended Advisory Board Committee, for their generous assistance and active participation in discussions, carried out either during live sessions as well as over the Internet, and (c) the anonymous ASIS&T 2006 referee for the most constructive comments made. Special thanks are due to ISI-Thomson Scientific (http://www.isinet.com/) for making their citation database available in order to be utilized along the lines of C-CAP.

REFERENCES

Bollen, J. , and Rodriguez, M.A., Van de Sompel H. (2006). Journal Status. Retrieved on June 19, 2006 from: http://arxiv.org/abs/cs.DL/0601030

Brin, S., and Page, L. (1998). The Anatomy of Large-Scale Hypertextual Web Search Engine. Retrieved on June 19, 2006 from:
http://www.public.asu.edu/~ychen127/cse591f05/anatomy.pdf

Coleman, A. (2006). Assessing the Value of a Journal Beyond the Impact Factor. Journal of Education for Library and Information Science. Submitted to Journal of the American Society for Information Science & Technology. Retrieved on June 19, 2006 from: http://dlist.sir.arizona.edu/1030/

Dervos, D.A., and Kalkanis, T. (2005). cc-IFF: A Cascading Citations Impact Factor Framework for the Automatic Ranking of Research Publications. Proceedings of the 3rd IEEE International Workshop on Intelligent Data Acquisition and Advanced Computer Systems: Technology and Applications (IDAACS), p. 668-673, Sofia, Bulgaria, 5-7 September, 2005. Postprint version available from DLIST. Retrieved on June 19, 2006 from: http://dlist.sir.arizona.edu/1105/

Dunham, M.H. (2003). *Data Mining: Introductory and Advanced Topics*, Prentice Hall.

Garfield, E. (1955). Citation Indexes to Science: a New Dimension in Documentation through Association of Ideas. *Science 122*(3159), 108-111. Retrieved on June 19, 2006 from: http://www.garfield.library.upenn.edu/essays/v6p468y1983.pdf

Garfield, E.. (1999). Journal Impact Factor: a brief review. *Canadian Medical Association Journal*, *161*(8), Retrieved on June 19, 2006 from: http://www.garfield.library.upenn.edu/papers/journalimpactCMAJ1999.pdf

Garfield, E. (2005). The Agony and the Ecstasy - The History and the Meaning of the Journal Impact Factor. Presented at the International Congress on Peer Review and Biomedical Publication, Chicago, USA, September 16, 2005. Retrieved on June 19, 2006 from: http://www.garfield.library.upenn.edu/papers/jifchicago2005.pdf

Giles, C.L., Bollacker, K., and Lawrence, S. (1998). CiteSeer: An Automatic Citation Indexing System, Digital Libraries 98 -Third ACM Conference on Digital Libraries Proceedings, 89-98

Glänzel, W., and Moed, H. F. (2002), Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193.

Glänzel, W., Thijs, B., and Schlemmer, B. (2004). A Bibliometric Approach to the Role of Author Self-Citations in Scientific Communication. *Scientometrics*, *59*(1), 63-77.

Hitchcock, S. (2003). Core Metalist of Open Access Eprint Archives: The Genesis of Institutional Archives and Independent Services, *ARL Bimonthly Report 227*, Retrieved on June 19, 2006 from: http://www.arl.org/newsltr/227/metalist.html. Updated version retrieved on June 19, 2008 from: http://opcit.eprints.org/explorearchives.shtml

Hoeffel, C. (1998). Journal Impact Factors [letter]. *Allergy 53*, 1225-1225.

Kleijnen J.P.C. and Van Groenendaal, W. (2000). Measuring the quality of publications: new methodology and case study. *Information Processing and Management*, *36*(4), 551-570

Moed, H.F., Burger, W.J.M., Frankfort, J.G., and Van Raan, A.F.J. (1985). The application of bibliometric indicators: important field-and time-dependent factors to be considered. *Scientometrics 8*(3-4), 177-203.

Moed, H.F. (2005). Citation Analysis of scientific journals and journal impact measures. *Current Science 89*(12),1990-1996.

Pinski, G., and Narin, F. (1976). Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics, *Information Processing and Management*, *12*(5), 297-312.

Smith, L. (1981). Citation Analysis. *Library Trends 30*, 83-106.

Snyder, H. and Bonzi, S. (1998). Patterns of Self-citations across disciplines (1980-1989). *Journal of Information Science, 24*(6)*,* 431-435.

Tomer, C. (1986). A statistical assessment of two measures of citation: the impact factor and the immediacy index, *Information Processing and Management*, *22*(3): 251-258.

Weertman, N. (2006). The New Scopus Citation Tracker. *Inside Scopus*, March 2006 Issue, 4-6. Retrieved on June 19, 2006 from: http://www.info.scopus.com/is