ISSI Newsletter

QUARTERLY E-NEWSLETTER OF THE INTERNATIONAL SOCIETY FOR SCIENTOMETRICS AND INFORMETRICS
ISSN 1998-5460

#40 / VOLUME 10 NUMBER 4
DECEMBER 2014

CONTENTS

EDITORIAL

ISSI NEWSLETTER COMING OF AGE

WOLFGANG GLÄNZEL **SARAH HEEFFER**

Centre for R&D Monitoring (ECOOM) & Dept. MSI, KU Leuven, Belgium

This is already the third occasion to write a résumé about the newsletter that has been launched ten years ago. The idea was to inform ISSI members in a fast and efficient manner about important events and affairs relevant to the community, in general, and to the society, in particular. The e-zine was preceded by information that has been published in the Journal Scientometrics and through other relevant communication channels rather sporadically. The new strategy was also reflected by the policies of the ISSI Board: The most recent number was reserved for society members while older issues were opened to the broad public. And this strategy, the periodical and fast electronic publication and open-access policies for the archive, proved to

ISSI e-Newsletter (ISSN 1998-5460) is published by ISSI (http://www.issi-society.info/).
Contributors to the newsletter should contact the editorial board by e-mail.

- **Wolfgang Glänzel,** *Editor-in-Chief:* wolfgang.glanzel[at]econ.kuleuven.be
- **Balázs Schlemmer,** *Technical Editor*: balazs.schlemmer[at]gmail.com
- **Judit Bar-Ilan**: barilaj[at]mail.biu.ac.il
- **Sujit Bhattacharya**: sujit_academic[at]yahoo.com
- **María Bordons**: mbordons[at]cindoc.csic.es
- **Jacqueline Leta**: jleta[at]bioqmed.ufrj.br
- **Olle Persson**: olle.persson[at]soc.umu.se
- **Ronald Rousseau**: ronald.rousseau[at]khbo.be
- **Dietmar Wolfram**: dwolfram[at]uwm.edu

ISSI — international society for scientometrics and informetrics

become a success story. This has already been pointed out in the two previous résumés (Glänzel, 2008; Glänzel and Heeffer, 2012). However, the recipe for success is perhaps also due to the unique mixture of information and entertainment combined with a number of pre-prints and research notes on topical and even "hot" issues. As a results, the robust foundation of the Newsletter is formed by three pillars: the up-to-date information about upcoming events and society affairs, the meeting reports and interviews, and, finally, the research notes, discussion papers and book reviews. Some of these pieces proved to be real breakthroughs, which is not only reflected by their citation impact but also by the media attention and publicity these publications have earned. Being veritable bibliometricians we have compiled statistics and discussed some of these papers already in 2012. Now time has come to repeat and update this exercise for checking sustainability of this impact. Since January 2005 four book reviews and 92 articles and short communications have been published, that is, 24 new research notes since the last résumé. We have searched for citations received by them till 17 December 2014 in three databases: Thomson Reuters Web of Science (WoS), Elsevier's SCOPUS and Google Scholar (GS) using Harzing's 'Publish or Perish' tool (Harzing, 2007). First we have updated the citation-frequency plot of the previous editorial by Glänzel and Heeffer (2012). The citation rates of all papers, that have received at least ten citations in one of these databases, are shown in Figure 1. Among these we find several frequently cited papers, even highly cited papers – at least as compared with the standard in our field. Egghe's paper on the *g-index*, which was already mentioned as highly cited in 2012, proved to be a true citation classic: The number of citations this paper attracted increased by 50% since 2012, from 81 to 122 according to the WoS data. In the mirror of Google Scholar this increase is even more dramatic: 147 citations in 2012 are contrasted by 258 just three years later. WoS and Scopus reflect similar citation rates with little deviations from each other – in both directions. Therefore we will mainly focus on the Web of Science and Google Scholar. In this context we have to mention that the sometimes extremely high citation rates in GS have to be taken with a pinch of salt as the set of citing documents might contain some duplicates or dead links.

A closer look at those pieces, that have received at least 20 citations each, uncovers what was considered as 'hot' in our field, when the papers appeared. As already mentioned, the most cited paper, was concerned with Hirsch-type indicators. The same applies to Kosmulski's (2006) article with 82 citations according to the WoS (still 50 in 2012) and Jin (2007) with 55 WoS citations in 2014 (40 three years ago). GS reports 165 citations

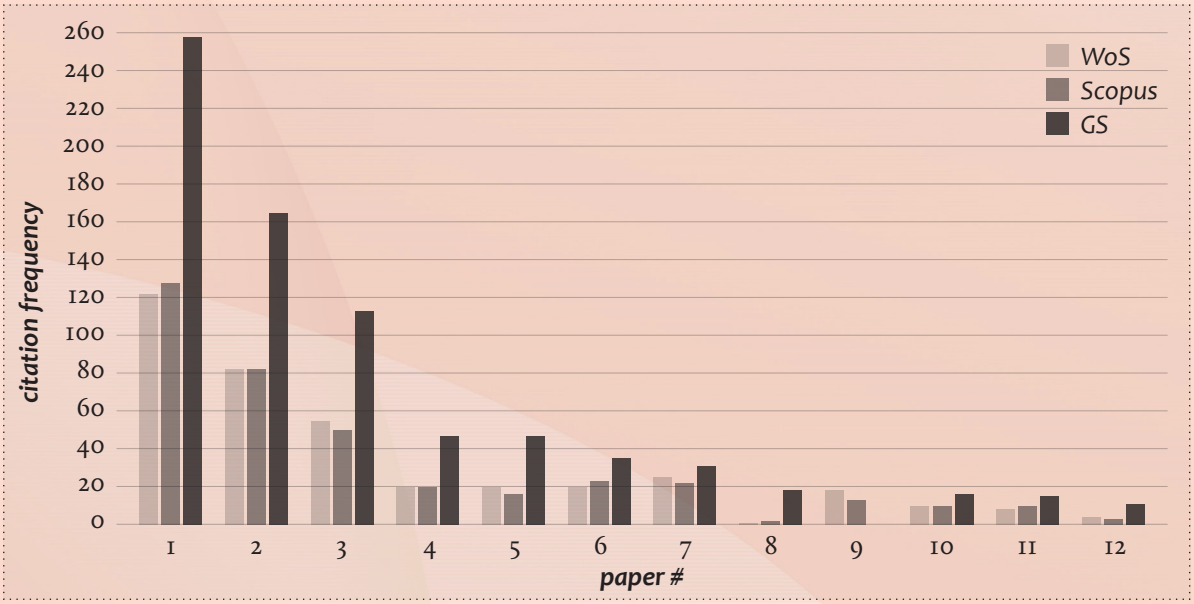


Figure 1 *The ISSI Newsletter in the mirror of citation indices (2005-2014)*

for Kosmulski and 113 for Jin, respectively, by the end of this year. The pieces by Bar-Ilan (2006) and Glänzel and Persson (2005) were cited somewhat less frequently (25 times in 2014/23 times in 2012 for Bar-Ilan; 22/20 cites for Glänzel and Persson according to the WoS and 31/24 and 47/34 times, respectively, according to GS), but were devoted to the same topic. In our previous editorial we mentioned already that the newsletter was among the first periodicals to react on the new performance indicator and that this initiative was rewarded by the community through its attention and, of course, more measurably through citations.

Two other papers, that have received 20 WoS citations each till December 2014, are concerned with completely different topics, but have nevertheless attracted much attention. Sivertsen (2010) introduced a performance based funding model for the Higher Education Sector. This Norwegian model has attracted much attention and is followed by other regions and countries in Europe since. GS reports 35 citations till now.

The other paper with 20 WoS citations was published by Labbé (2010). GS recorded 47 citations by the end of this year. Using the bold experiment of generating bogus papers, Labbé addressed a clear warning to the community. He did not only point to gaps in the present computer-aided system of academic writing, reviewing and publishing but also showed in later publications (e.g., Labbé, 2012) how to detect duplicate and fake publications in the scientific literature. For his work, the article in the ISSI Newsletter was the initial spark. We already reported the media impact of his experiment three years ago and this impact is not only lasting but truly sustainable. Quite recently van Noorden (2014) reported a large withdrawal campaign by Springer and IEEE in Nature News and explicitly refers to Labbé's work in this context.

To summarise, there is evidence for the attention paid to the Newsletter and the sustainability of its impact. In some cases the impact was immediate, in others the Newsletter was rather used for publishing preliminary work or even just some interesting ideas the authors wished to share with the community. However, the number of authors is still limited as we have already noticed three years ago. And the editorial board is still somewhat overrepresented among the authors. And as last time we would like to encourage all readers of the Newsletter again to actively contribute to the continuation of the success of this project. We also take this opportunity to thank the readers of the Newsletter for their loyalty and the editorial staff and all contributors who have supported the Newsletter with their previous work as author and reviewer.

REFERENCES

Bar-Ilan, J. (2006), H-index for price medalists revisited. *ISSI Newsletter*, 2 (1), 3-5.

Egghe, L. (2006), An improvement of the h-index: the g-index. *ISSI Newsletter*, (2) 1, 8-9.

Glänzel, W., Persson, O. (2005), H-index for price medalists. *ISSI Newsletter*, 1 (4), 15-18.

Glänzel, W. (2008), Three years of ISSI Newsletter. *ISSI Newsletter*, 4 (1), 1-2.

Glänzel, W., Heeffer S. (2012), The seven-year itch? *ISSI Newsletter*, 8 (1), 1-3.

Harzing, A.W. (2007), *Publish or Perish*. Available from http://www.harzing.com/pop.htm.

Kosmulski, M. (2006), A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, 2 (3), 4-6.

Labbé, C. (2010), Ike Antkare one of the great stars in the scientific firmament. *ISSI Newsletter*, 6 (2), 48-52.

Sivertsen, G. (2010), A performance indicator based on complete data. *ISSI Newsletter*, 6 (1), 22-28.

van Noorden, R. (2014), Publishers withdraw more than 120 gibberish papers. *Nature*, doi:10.1038/nature.2014.14763.

REPORT ON THE 19th INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS

3–5 SEPTEMBER 2014, LEIDEN, THE NETHERLANDS

ED NOYONS
CWTS, The Netherlands

PAUL WOUTERS
CWTS, The Netherlands

INTRODUCTION

The Science & Technology Indicators (STI) conference was held for the first time in 1988 in Leiden and returns here traditionally every 4 years. Since 2005 the STI conference hosts also the ENID conference, making it an annual meeting of researchers in the fields of research evaluation metrics and their users. The 19th edition of the STI was held in Leiden, the Netherlands 3-5 Sept 2014 (sti2014.cwts.nl). The Centre for Science and Technology Studies (CWTS), Leiden University, organized the conference and welcomed almost 250 participants. Of the Leiden editions this was the largest conference and coincided with the 25th anniversary of CWTS. For this special occasion, a team of actor-journalists, De Waan, was active during the breaks to lighten the atmosphere as well as to ironically reflect on topics and discussions in the workshops and plenaries. They launched twice a day

an edition of their 'Daily Issues', which was received very well by the participants.

The motto or theme of this year's edition was 'Context counts: pathways to master big and little data'. The aim was to give more room for interpretation of indicators and data and to provide a bigger platform for the use of the indicators and measurement in general. This reflects the surge in the uptake and use of science & technology indicators at universities, research institutes and in public policy contexts.

PROGRAM AND EVENT

Eventually 125 papers or posters were submitted. These submissions were reviewed by a hundred members of the scientific committee. The large amount of reviewers allowed us to have each submission to be judged by three persons. No member had to review more than 5 papers. On the basis of the judgments of the scientific committee we were able to host 52 short papers, 18 research in progress papers, and 30 posters. Over 20 submissions were rejected or withdrawn. Parallel to the traditional submissions, people were able to submit proposals for special sessions. In total 9 proposals were submitted. Five of them were admitted to the program.

The theme of this edition of the STI was intended to broaden the scope of the conference. Context of the indicators and data refers to the environment in which these indicators are used as well as to their interpretation. We were able to organize some 20 sessions with a variety of topics. Some of them could be based on previous editions but there were also new topics, e.g., behavior of scientists and careers & trajectories.

The poster session deserve special mention in this report. Conference participants devoted a lot of attention to the preparation of beautiful and clear posters. The Award for Best Poster was won by Sabrina Petersohn. The award entails a research stay at CWTS.

The special sessions (mainly panel discussions) provided an interesting platform for discussion and development. And although the plenary panels consume much of the program, most of them were well received. During one of the panels, the foundation for metrics standards was laid for the development of principles of good evaluation practices, which will be further developed by the community. The ENID association decided that it would disseminate a draft declaration on the use of bibliometric indicators among its members as the basis for a public ENID document.

An important contribution to the broadening of the scope was done by the keynote speakers. The opening address of Peter Dahler-Larsen set the stage and was referred to many times in presentations and discussions. In addition, the keynotes by Diana Hicks and Eppo Bruins further deepened the conference themes.

CONCLUSION

The 2014 edition of the STI conference was a huge success in many ways. We have never had so many participants in Leiden. Moreover, important steps were taken to broaden the scope of the conference as well as to develop principles and professional standards within the community. And thanks to the effort of the local organization, participants enjoyed the event very much.

The next edition of the STI conference will be held in Lugano, 2-4 Sept 2015. More information at http://www.sti2015.usi.ch.

TECHNICAL WORKSHOP ON BIBLIOMETRIC INDICATORS

WORKSHOP REPORT

LUDO WALTMAN
CWTS,
The Netherlands

PAUL WOUTERS
CWTS,
The Netherlands

NEES JAN VAN ECK
CWTS,
The Netherlands

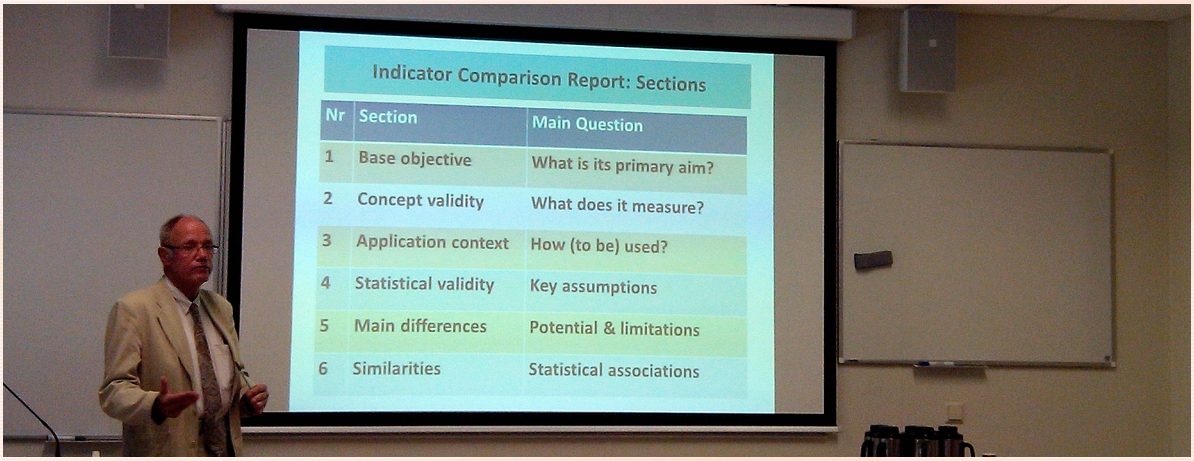
TINA NANE
CWTS,
The Netherlands

INTRODUCTION

On September 2nd, 2014, the Centre for Science and Technology Studies (CWTS) of Leiden University organized a one-day technical workshop on bibliometric indicators. The workshop took place in Leiden, the Netherlands immediately before the STI conference. In order to have a well-focused discussion, it was decided to have only a limited number of participants in the workshop. There were about 15 invited participants, as well as about 10 participants from CWTS. As organizers of the workshop, we would like to present a brief report of the discussions that took place during the workshop.

TOPICS OF DISCUSSION

The first topic was about the strengths and weaknesses of different bibliometric indicators. Talks were given by Wolfgang Glänzel and Ludo Waltman. Wolfgang emphasized the importance of thinking about indicators not only from a scientometric point of view, but also from a mathematical one. He presented a systematic perspective on the issues to be taken into consideration in the design of indicators, pointing out for instance the problem of the large confidence intervals of the h-index and the limitations of composite indicators. Ludo discussed the problem of the ranking inconsistency of the h-index and the problem of the extreme

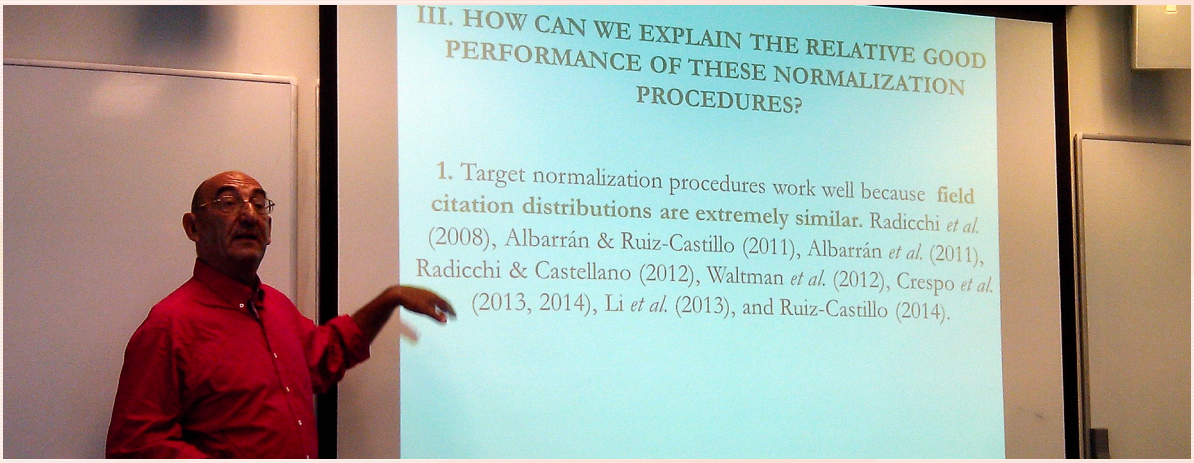


sensitivity of average-based indicators to 'outliers', arguing that percentile-based indicators appear to offer the most satisfactory measures of citation impact.

The second topic was about the normalization of citation impact indicators, with talks by Michel Zitt and Javier Ruiz-Castillo. Michel discussed the three basic approaches to normalization: The cited-side approach, the citing-side approach, and the recursive network-based approach. He then offered his perspective on the properties of the three approaches, emphasizing in particular that the cited-side approach provides a kind of total normalization while the citing-side approach provides a more partial normalization. Javier focused on the problem of comparing the performance of different normalization approaches. He paid special attention to the role played by field classification systems both in the application and in the comparison of normalization approaches. He also stressed the importance of the similarity in citation distributions in analyzing normalization approaches.

The third topic covered in the workshop was about country-level and institutional-level analyses. Jonathan Adams focused on the country level and emphasized the importance of choosing citation windows in a proper way. He showed how the performance of a country may look quite differently depending on the way in which citation windows are chosen. Especially longitudinal analyses are affected by this issue. Nees Jan van Eck, focusing mainly on institutional-level analyses, raised the issue of different counting approaches for handling co-authored publications, in particular the full and the fractional counting approach. Nees Jan argued that proper field normalization is not possible using full counting but can be achieved using fractional counting. Full counting will benefit institutions in fields in which there is a lot of collaboration and in which collaboration is strongly correlated with citations. In particular medical institutions are advantaged by the use of full counting.

In the afternoon, the first topic of discussion was statistical inference in bibliometric analyses. Presentations were given by Wolfgang Glänzel, Jesper Schneider, and Tina Nane. Wolfgang discussed the importance of stochastic models in scientometric analyses. He underlined that even though scientometric distributions tend to be strongly skewed, indicators derived from these distributions usually do have approximately normal distributions. This was illustrated for the h-index. Jesper presented an argument in favor of Bayesian rather than frequentist inference. He also criticized the superpopulation idea that is sometimes used to justify statistical inference in situations in which the entire population rather than just a sample is available. Tina distinguished between descriptive and inferential analyses, arguing that inferential analyses may be justified using a superpopulation argument. She also showed the relationship between



confidence intervals and hypothesis tests and the use of bootstrapping techniques to analyze this relationship.

Journal indicators were the final topic of the workshop, with presentations by Henk Moed, Vicente Guerrero Bote, Vincent Larivière, and Ismael Rafols. Henk introduced the idea of indicator comparison reports for making careful comparisons between different journal indicators. He illustrated this idea by presenting a comparison between the original and the modified version of the SNIP indicator. Vicente suggested a new type of journal indicator, focusing not on the scientific impact of journals but on their role in technology transfer. The proposed indicator, referred to as the 'technological factor', is based on citations given in patents to scientific journals. Vincent strongly argued against the use of journal indicators in the evaluation of researchers, institutions, and countries. He pointed to the high skewness of citation distributions, implying that the impact factor of a journal is only a weak predictor of the number of citations of individual publications in the journal. He also drew attention to the absence of a strong correlation between impact factors and rejection rates. Finally, Ismael proposed to look at journal indicators not only from a technical perspective but also from the perspective of indicators as social technology. He pointed to the effect indicators have on researchers' behavior and to the very limited influence scientific discussions on journal indicators have on the actual use of these indicators. He argued that journal indicators should be discussed not only from a technical supply perspective but also from a demand perspective focusing on the use of the indicators.

CONCLUSION

There was a lot of debate during the workshop and a very fruitful exchange of opinions. Although there was agreement on some issues, the participants in the workshop also turned out to have quite different perspectives on certain issues, especially on issues related to the basic properties good indicators should have, the issue of different counting methods, and the issue of proper ways to perform statistical inference. Given the wide variety in disciplinary backgrounds of the workshop participants (and of bibliometric researchers more generally), the complexity of the discussion was sometimes increased further by the use of different terminologies. The workshop should be seen as part of a broader discussion on the possibilities for standardization in scientometrics, a discussion that involves not only technical questions but also many questions that are related more to the use of indicators in all kinds of research assessment contexts.

The slides of the presentations given at the workshop can be downloaded from www.cwts.nl/pdf/workshop_bibliometric_indicators.zip.

COLLNET 2014

CONFERENCE REPORT

BERND MARKSCHEFFEL
Steering Committee Chair

HILDRUN KRETSCHMER
COLLNET Center, Gemany

COLLNET AND WIS HISTORY (WIS: WEBOMETRICS, INFORMETRICS, SCIENTOMETRICS)

COLLNET is a global interdisciplinary research network of scholars who are concerned to study aspects of collaboration in science and in technology (see COLLNET web site at: http://www.collnet.de/). This network of interdisciplinary scholars was established in January 2000 in Berlin with Hildrun Kretschmer as coordinator. Since that time there have been fourteen meetings: the first in Berlin, September 2000, the 2nd in New Delhi, February 2001 and the 3rd in Sydney (in association with the 8th ISSI Conference), July 2001. The 4th COLLNET Meeting took place on August 29th in 2003 in Beijing in conjunction with the 9th International ISSI Conference; the First International Workshop on Webometrics, Informetrics and Scientometrics (WIS) and 5th COLLNET Meeting in Roorkee, India, in March 2004. The 6th COLLNET Meeting took place in association with the 10th ISSI Conference in Stockholm, Sweden, in July 2005.

The Second International Workshop on Webometrics, Informetrics and Scientometrics (WIS) and 7th COLLNET Meeting was organized in Nancy, France, in May 2006.

The Third International Conference on WIS and Science and Society & Eighth COLLNET Meeting took place in New Delhi, India, in March 2007 (http://www.collnet-delhi.de), the Fourth International Conference on WIS & Ninth COLLNET Meeting in Berlin, Germany in July 2008 (http://www.collnet-berlin.de) and the Fifth International Conference on WIS & Tenth COLLNET Meeting in Dalian, China, in September 2009 (http://www.wiselab.cn/collnet-dalian/). The Sixth International Conference on WIS & Eleventh COLL-

NET Meeting took place in Mysore, India, in October 2010,the Seventh International Conference on WIS & Twelfth COLLNET Meeting in Istanbul, Turkey, in September 2011 (http://collnet.cs.bilgi.edu.tr/), the 8thInternational Conference on WIS & 13th COLLNET Meeting in Seoul, Korea, October, 2012, Seoul, Korea, http://collnet2012.ndsl.kr; the 9thInternational Conference on WIS &14thCOLLNET Meeting, August, 2013in Tartu, Estonia, http://www.etag.ee/international-research-cooperation/collnet-2013/?lang=en.

COLLNET MEETING AND INTERNATIONAL CONFERENCE IN 2014

The 10th International Conference on Webometrics, Informetrics and Scientometrics (WIS) & 15th COLLNET Meeting took place in Ilmenau, Germany, on 3-5 September 2014. This joint meeting was organised under the auspices of the international organisation COLLNET and by the TU Ilmenau, Ilmenau, Germany.

SCOPE AND ORGANIZING COMMITTEE

The broad focus of the conference is on collaboration and communication in science and technology; science policy; quantitative aspects of science of science; and combination and integration of qualitative and quantitative approaches in study of scientific practices.

The conference thus aims to contribute to evidence-based and informed knowledge about scientific research and practices which in turn may further provide input to institutional, regional, national and international research and innovation policy making.

- *General Chair:*
  Hildrun Kretschmer (Germany, China)
- *Steering Committee Chair:*
  Bernd Markscheffel (Germany)
- *Programme Committee:*
  COLLNET Members
  http://www.collnet.de/ and
- *Local Programme Committee:*
  Bernd Markscheffel
  Daniel Fischer
  Bastian Eine
  Daniela Büttner

- *Regional Chairs:*
  Valentina Markusova (Russia)
  Liang Liming (China)
  Ramesh Kundra (India)
  + N.K. Wadhwa (India)
  + Divya Srivastava (India)
  + Sujit Bhattacharya (India)
  + P.K. Jain (India)
  Farideh Osareh (Iran)

PARTICIPANTS AND PROGRAMME:

The 10th International Conference attracted more than 80 participants from 20 countries:

- Europe (10): Belarus, Denmark, France, Germany, Hungary, Russia, Spain, The Netherlands, Turkey, UK
- Asia (8): China, India, Indonesia, Iran, Japan, Korea, Sri Lanka, Taiwan
- America (2): Canada, USA

At the conference, five keynote lectures were presented and two invited papers as well as about 70 oral and poster presentations.

Keynote Speakers:

- Eugene Garfield (USA) & Alexander Pudovkin (Russia)
- Liang Liming (China), Zhen Zhong (China) & Ronald Rousseau (Belgium)
- Weiping Yue (China)
- Sujit Bhattacharya (India):
- Hildrun Kretschmer (Germany, China) & Theo Kretschmer (Germany)

Invited Speakers:

- I. K. Ravichandra Rao (India)
- P.K. Jain (India)

The Keynote speaker **Eugene Garfield** and his co-author **Alexander Pudovkin** have presented the talk entitled "Journal Impact Factor Reflects Citedness of the Majority of the Journal Papers". The speakers have mentioned the literature on Journal Impact Factors (JIF) is quite rich with the assertion that one of the main drawbacks of the JIF is

its presumptive dependence on only a few highly cited papers published in the journal. But in their paper the speakers wished to test the validity of the claim that the JIF reflects the citedness of the majority of a journal's papers or, put it another way to disprove the widely reported myth that the JIF depends on only a few highly cited papers.

The talk entitled "Uncited Papers, Uncited Authors and Uncited Topics" by the Keynote Speaker **Liang Liming** and her co-authors **Zhen Zhong** and **Ronald Rousseau** was presented by Weiping Yue. The Keynote speaker and her co-authors have studied the question what are the similarities and differences of the bibliometric characteristics between 1) uncited and cited papers; 2) uncited and cited authors; 3) uncited and cited topics. The authors explored and discussed these problems: How does the group of uncited authors look like? Does produc-

tive and highly cited author publish uncited paper(s)? Why does a research topic become an uncited topic?

Weiping Yue, the Keynote speaker on the topic "A Scientometric Study on Collaboration between Academia and Industry" has presented the development of this special collaboration in China. This collaboration covers various forms of engagement, i.e. joint research contract research, patent transfer and technology transfer. A scientometric analysis was applied to data collected from leading universities and companies in China, who are ranked as top entities in terms of total number of inventions in the white paper of Research & innovation performance of the G20.

The talk on the topic "International Collaboration: Changing the Global Landscape of Science in the 21st Century" was held by the Keynote speaker **Sujit Bhattacharya**.

In the last decades new concepts have emerged in order to understand pattern formation in complex networks of interactions. The Keynote speaker **Hildrun Kretschmer** and her co-author **Theo Kretschmer** have presented three-dimensional visualization and animation of emerging patterns by the process of self-organization in collaboration networks. The corresponding well-ordered 3-D computer graphs are totally rotatable around and their shapes are visible in the space from all possible points of view. The objectives of the paper are the description of several methods for three-dimensional modelling and animation and the application of these methods on two co-authorship networks selected for demonstration of varying 3-D graph images.

I. K. Ravichandra Rao and his co-author **K. S. Raghavan** (Invited speakers) have explained the results of the study entitled "Seven years of the 'COLLNET Journal of Scientometrics and Information Management' " (2007-2013). In this connection Ruppesh Kr Gupta has delivered the publishing report, August 2014, about the 'COLLNET Journal of Scientometrics and Information Management' by Taylor & Francis.

The event of COLLNET 2015: 11th International Conference on Webometrics, Informetrics and Scientometrics (WIS) & 16th COLLNET Meeting, November 2015, India, was introduced by the Invited speaker **P.K. Jain**.

Oral and Poster Presentations:
The variety of topics given by the keynotes and invited papers is also mirrored by the about 70 oral and poster contributions. The parallel sessions were mainly focused on the following topics: Theoretical Approaches and Methodology, Citations and Evaluation, Collaboration and Communication, Webometrics, Informetrics and Scientometrics, Gender and Network Analysis, Technology and Engineering Studies, National Oriented Studies and Miscellaneous.

Besides the Proceedings of the WIS & COLLNET Meeting provided to the participants of the conference in Ilmenau, a selection of keynote and contributed papers will be published in 2015 in dedicated issues of the international *COLLNET Journal of Scientometrics and Information Management* (Taylor & Francis Group, UK, co-published by TARU Publications, India)

The journal publishes original research papers. The journal is available in print and online at www.tandfonline.com/tsim.

CONFERENCE VENUE: ILMENAU UNIVERSITY OF TECHNOLOGY

Ilmenau University of Technology is the only university in the federal state of Thuringia with the title "Technische Universität". Research and education is focused on engineering with strong links to economics and natural sciences. It was founded in 1894 and has a total of 5 academic faculties and

about 7,200 students. Personal care for students from professors, tutors and student mentors; a campus with modern buildings; only short distances apart; a variety of social activities and social support; many student associations as well as diverse cultural and sports activities are among the distinguishing features of TU Ilmenau.

Research and education at Technische Universität Ilmenau is focused on engineering with strong links to economics and natural sciences. TU Ilmenau has a long tradition in information science as well as in science and technology education. Over 7000 students studying for Bachelor's and/or Master's degrees in which the subjects tend to be drawn from a number of disciplines within the overall groups of Engineering, Mathematics with Science, and Economics with Social Studies. The Institute of Business Informatics of the Faculty of Economic Science and Media as organiser takes an important role in scholarly communication for improving co-work and collaboration among researchers and practitioners worldwide. It maintains close relationships with various research institutes and enterprises.

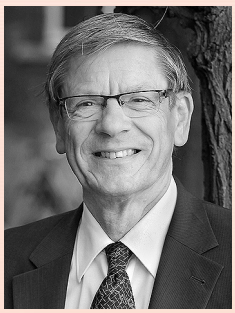
COLLNET MEETING AND INTERNATIONAL CONFERENCE IN 2015

The following event of this conference series will be organised in the coming year. The 11th International Conference on Webometrics, Informetrics and Scientometrics (WIS) & 16th COLLNET Meeting will be held in Delhi, India 26-28 November 2015, organized by the Society for Library Professionals (SLP) in association with Asian Chapter, Special Libraries Association (SLA) and Institute of Economic Growth (IEG), University of Delhi, India. http: www.slp.org.in/collnet2015

AN APPROACH FOR EFFICIENT ONLINE IDENTIFICATION OF THE TOP-*K* PERCENT MOST CITED DOCUMENTS IN LARGE SETS OF WEB OF SCIENCE DOCUMENTS

PER AHLGREN
School of Education and Communication in Engineering Sciences (ECE), KTH Royal Institute of Technology, Stockholm, Sweden
perahl@kth.se

OLLE PERSSON
Department of Sociology, Inforsk, Umeå University, Umeå, Sweden
olle.persson@soc.umu.se



RONALD ROUSSEAU
Institute for Education and Information Sciences, IBW, University of Antwerp, Antwerp, Belgium;
KU Leuven, Leuven, Belgium
ronald.rousseau@kuleuven.be

ABSTRACT:
The citation indices of the Web of Science play an important role as data sources in evaluative bibliometrics. Normally, the Web of Science is accessed online via the interface provided by Thomson Reuters. Due to the 500-documents-per-time export restriction, it is practically infeasible to export sets that may contain tens of thousands of documents. Moreover, at most 100,000 retrieved documents are visible in the interface. In this work, we present an approach for efficient online identification of the top-k percent most cited documents in large sets of Web of Science documents. We also report a small study, the main purpose of which is to illustrate the presented approach, of the performance of four European countries–Denmark, the Netherlands, Sweden and Switzerland–regarding the top-1 percent most cited articles in the field *Biomedical and Health Sciences*.

KEYWORDS:
biomedical and health sciences; citation statistics; country performance; percentiles; Web of Science

1. INTRODUCTION

The citation indices of the Web of Science (WoS) play an important role as data sources for evaluative bibliometrics. This is the case not only for gauging research efforts, but also for analyses performed by library or administrative staff at higher education institutions (HEIs) and by staff at various governmental units, such as research councils. In Sweden, for instance, several HEIs have library staff that analyses the research of different units of the institutions using bibliometric methods based on WoS data. An example of such a HEI is Stockholm University (see http://www.sub.su.se/publish/bibliometrics.aspx). Similarly, the Swedish Research Council uses the WoS to perform annual analyses of the citation impact and publication production of most HEIs.

When performing such analyses it is preferable to have access to a version of the WoS adjusted to bibliometric needs. An example of a unit having such a bibliometric version is CWTS, Leiden University, the Netherlands[1]. However, most colleagues performing research evaluations only have access to the online version of the WoS as provided by Thomson Reuters. Considered from a bibliometric point of view, this interface has several drawbacks. One of these is the 500-documents-per-time export restriction. Due to this restriction, it is practically infeasible to export sets that may contain tens of thousands of documents. Moreover, at most 100,000 retrieved documents are visible in the interface.

Nowadays percentiles (quantiles in general) are frequently used in citation statistics. To measure the proportion of documents of a given unit of analysis, like an institution, that belong to the top-k percent most cited (with respect to world-wide citation distributions of similar documents in terms of field, publication year and document type) complements the measurement of an average citation impact of the unit. For the latter, the mean (field) normalized citation score, used in the CWTS Leiden Ranking 2014[2], is a common indicator choice (Waltman et al. 2011a; Waltman et al. 2011b). For the former, common values of k are 1, 5 and 10 (e.g., CWTS Leiden Ranking 2014; Karlsson and Persson 2012; Visser and Nederhof 2011).

The aim of this article is twofold: (a) to put forward an approach for efficient online identification of the top-k percent most cited documents in large sets of WoS documents (this set of documents will be referred to as the set of target documents), and (b) to present an application of this approach in a small study of the performance of four European countries with respect to the top-1 percent most cited documents in the field *Biomedical and Health Sciences*.

The remainder of this paper is organized as follows. In the next section, the approach for identification of the top-k percent most cited documents in large sets of WoS documents is described. The study referred to under (b) above is reported in the following section, whereas concluding remarks are given in the last section.

2. IDENTIFICATION OF THE TOP-k PERCENT MOST CITED DOCUMENTS

It is well-known that, in the discrete case, the definition of a quantile is not unequivocal (Hyndman & Fan, 1996). In this study we cover four definitions of the p^{th} ($0 < p < 100$) percentile. Three of these definitions correspond to the three quantile algorithms given in Hyndman and Fan (1996) in the Section *Discontinuous Functions* (algorithms 1, 2 and 3). The fourth definition, proposed already by Hazen (1914) makes use of linear interpolation, and corresponds to algorithm 5 in the Section *Piecewise Linear Continuous Functions* in the same paper by Hyndman and Fan. This definition has been

1 http://www.cwts.nl/

2 http://www.leidenranking.com/

used recently in (Bornmann, Leydesdorff and Mutz 2013; Bornmann, Leydesdorff and Wang 2013). Moreover, we exclude sets with no or an extremely small number of elements. Assume that the observations, $x_1, \ldots x_n$, are placed in ascending order. According to Definition A (corresponding to algorithm 1), the p^{th} percentile is the smallest x such that $F(x) \times 100 \geq p$, where F is the cumulative distribution function for the considered empirical distribution, in other words: $F(x)$ is the share of observations $\leq x$. Definition B (corresponding to algorithm 2) is similar to Definition A. The only difference is that when $p / 100 \times n$ is an integer m, the p^{th} percentile is equal, not to the m^{th} observation, x_m, but to the mean of observations x_m and x_{m+1}. Definition C (corresponding to algorithm 3) uses a nearest integer approach. If $p / 100 \times n$ cannot be written in the form $j.5$, where j is an integer, $p / 100 \times n$ is rounded to its nearest integer, m, and the p^{th} percentile is equal to x_m. If $p / 100 \times n$ can be written in the form $j.5$, $p / 100 \times n$ is rounded to j if j is even, otherwise to $j + 1$. In the first case, the p^{th} percentile is equal to x_j, in the second case to x_{j+1}. Definition D (corresponding to algorithm 5) uses linear interpolation. If $p/100 \times n + 0.5$ is a non-integer y, then there is an integer j such that $j < y < j + 1$, and the p^{th} percentile is equal to $x_j + (y - j)(x_{j+1} - x_j)$. Thus, in this case the p^{th} percentile is taken to be the interpolated value between the j^{th} observation and observation $j + 1$. If $p / 100 \times n + 0.5$ is an integer m, then the p^{th} percentile is equal to x_m.

The four quantile algorithms corresponding to the four definitions are implemented, for instance, in R, a free software environment for statistical computing and graphics ("The R project for statistical computing" 2014).

Assume that a WoS query has retrieved a large set of documents, say S. Let n_S be the number of documents in S. Go to the Results page in the WoS, and select Show 10 per page (which is the default). First we recall the procedure to visit the page containing a document ranked i (where, for the moment, the ranking criterion does not matter). Let r be the smallest integer $\geq i / 10$. Then go to page r, where the document with rank i is located (one may choose this page in the Page field situated on the same line as the number of retrieved documents). If, however, you now change the ranking criterion, for instance you want to sort the retrieved documents descending by citation values (Sort by: Times Cited – highest to lowest), then you do not stay at the page of the document ranked i (now according to the new ranking criterion), given that i is greater than 10 (as the page with the 10 most highly ranked, according to the new ranking criterion, documents are shown). Hence, you'd better first rank documents according to the criterion you are interested in.

We proceed to identify the top-k percent, where $k = (1, 2, \ldots, 99)$, most cited documents in S, our target documents.[3] We, moreover, consider as target documents the documents with citation values strictly larger than the $p^{th} = (100 - k)^{th}$ percentile in the citation distribution for the documents in S. Let $v = (c_1, \ldots, c_{n_S})$ be this distribution, where the values are ranked in descending order (note the change in ranking order). Let x be k percent of n_S, i.e., $x = k / 100 \times n_S$. Let m be the smallest integer $\geq x$. We first treat the scenario in which Definitions A or B are used, considering two cases depending on $x = m$ or $x \neq m$.

a) The case $x = m$. Go to the page of document d with rank $m + 1$ (type the page number in the Page field on the Results page). For Definition A the number c_{m+1}, which is the citation value for document d, is equal to the $p^{th} = (100 - k)^{th}$ percentile in v. For Definition B the $p^{th} = (100 - k)^{th}$ percentile in v is $\frac{1}{2} c_m + \frac{1}{2} c_{m+1}$, the mean of the citation values for d and the document that immediately precedes d. In both cases the documents with a

3 Clearly, only low values of k, like 1, 5 and 10, are of interest in the present context.

citation value strictly larger than c_{m+1} are the target documents. Note that a citation value is strictly larger than c_{m+1} if and only if it is strictly larger than $\frac{1}{2}\, c_m + \frac{1}{2}\, c_{m+1}$.

b) The case $x \neq m$. In this case, there is no difference between Definitions A and B, thus, the percentile value is the same. Go to the page of document d with rank m. Now, the value c_m is equal to the $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v. The documents with a citation value $> c_m$ are the target documents. Thus for the definitions A and B the target documents are always the same.[4]

Next, we treat the scenario in which Definition C is used. Similar to Definitions A and B, we consider two cases.

a) The case $x = m$. Go to the page of document d with rank $m + 1$. The value c_{m+1} is equal to the $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v. The documents with a citation value $> c_{m+1}$ are the target documents. Obviously, when $x = m$, Definition C agrees with Definition A (cf. case (a) above).

b) The case $x \neq m$.
Let $y = (100 - k) / 100 \times n_S$.
We consider two subcases.

▸ Subcase (b1): x cannot be written in the form $j.5$, where j is an integer. Round y to its nearest integer, say r. If $r > y$, go to the page of document d with the rank m. Then the value c_m is equal to the $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v. The documents with a citation value $> c_m$ are the target documents. If $r < y$, go to the page of document d with the rank $m+1$. Then c_{m+1} is equal to the $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v. The documents with a citation value $> c_{m+1}$ are the target documents.

▸ Subcase (b2): x can be written in the form $j.5$. Then y can be written as $(n_S - j - 1).5$. Round y to $(n_S - j - 1)$ if $(n_S - j - 1)$ is even, otherwise to $(n_S - j - 1) + 1$. In the first case, go to the page of document d with rank $m + 1$. Now it holds that c_{m+1} is equal to the $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v. The documents with a citation value $> c_{m+1}$ are the target documents. In the second case, go to the page of document d with the rank m; then c_m is equal to the $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v and the documents cited strictly more than c_m are the target documents.

Finally, we treat the scenario in which Definition D is used. Again, we consider two cases.

a) The case $x = m$. Go to the page of document d with rank $m + 1$. The $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v is $\frac{1}{2}\, c_m + \frac{1}{2}\, c_{m+1}$, and the documents with a citation value strictly larger than c_{m+1} are the target documents.

b) The case $x \neq m$.
We consider two subcases.

▸ Subcase (b1): x cannot be written in the form $j.5$, where j is an integer. If $x - 0.5 < m - 1$, go to the page of document d with rank m. The $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v is equal to $c_m + ((m - 1) - (x - 0.5))(c_{m-1} - c_m)$, and the documents with a citation value strictly larger than c_m are the target documents. If $x - 0.5 > m - 1$, go to the page of document d with rank $m + 1$. The $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v is equal to $c_{m+1} + (m - (x - 0.5))(c_m - c_{m+1})$, and the documents with a citation value strictly larger than c_{m+1} are the target documents.

▸ Subcase (b2): x can be written in the form $j.5$. Go to the page of document d with rank m. The value c_m is equal to the $p^{\text{th}} = (100 - k)^{\text{th}}$ percentile in v. The documents with a citation value $> c_m$ are the target documents.

After the application of the described approach, the target documents may be exported to files for offline analysis. We conclude this section by giving an example,

4 However, if the target documents are taken to be all documents such that their citation values are greater than or equal to pth percentile, different sets of target documents might be obtained.

where the approach is applied to identify the top-1 (k = 1) percent most cited documents in a set of retrieved WoS documents. *Example 1.* We use the query given in Appendix 1 (the number of publication years is, however, restricted to one, namely the year 1991). This query (databases = SCI-EXPANDED, SSCI, A&HCI) retrieves 229,369 documents (actually of *article type*). Let S be the set of these documents. Thus, n_S = 229369.

1. Turn to the Results page in WoS, and select Show 10 per page. Rank the 229,369 documents in descending order by citation values.
2. Let $v = (c_1, \ldots, c_{229369})$ be the citation distribution for the documents in S, where the values are ranked in descending order. We want to find the 99th percentile of v.
3. $x = 1 / 100 \times 229369 = 2293.69$. Then $m = 2294$, and hence $x \neq m$.
4. Definitions A, B, C and D
   a) *Definitions A and B.*
      Since $x \neq m$, case (b) is applicable. Go to the page of document d with rank 2294. The smallest integer greater than or equal to 2294 / 10 is 230. Thus, go to page 230. At the time of writing[5], document d is *Differential release of amino-acids, neuropeptides, and catecholamines from isolated nerve-terminals*, authored by Verhage et al.; its citation value c_{2294} = 256, the 99th percentile of the citation distribution v. Go back to the list of documents in order to select the target documents. These are all articles which are cited at least 257 times.
   b) *Definition C.*
      Since $x \neq m$, case (b) is applicable. $y = (100–1) / 100 \times 229369 = 227075.3$. x = 2293.69 cannot be written in the form j.5, where j is an integer, so subcase (b1) is applicable. The integer nearest to y is 227075, which is less than y. Go to the page of document d with the rank $m + 1$ = 2295 (page 230). At the time of writing, document d is *cDNA cloning of a myosin heavy-chain isoform in embryonic smooth-muscle and its expression during vascular development and in arteriosclerosis*, authored by Kuroo et al.; its citation value c_{2295} = 256, the 99th percentile of the citation distribution v. Now, return to the list of documents in order to select the target documents. These are all documents that are cited at least 257 times.
   c) *Definition D.*
      Since $x \neq m$, case (b) is applicable. x = 2293.69 cannot be written on the form j.5, where j is an integer. Therefore, subcase (b1) is applicable. Since 2293.69 – 0.5 > 2293, go to the page of document d with rank $m + 1$ = 2295. The 99th percentile in v is equal to
      $c_{m+1} + (m - (x - 0.5))(c_m - c_{m+1}) =$
      $256 + (2294 - (2293.69 - 0.5))(256 - 256) =$
      $256 + (2294 - (2293.69 - 0.5))(0) =$
      256
      Return to the list of documents in order to select the target documents. These are all documents that are cited at least 257 times.

In this example, the four percentile definitions give rise to the same set of target documents. However, Definition 3, as well as Definition 4, might give rise to a target document set that is distinct from the corresponding set for Definitions 1 and 2. Moreover, Definition 4 might give rise to a target document set that is distinct from the corresponding set for Definition 3.

3. PERFORMANCE OF FOUR EUROPEAN COUNTRIES REGARDING THE TOP-1 PERCENT MOST CITED ARTICLES IN THE FIELD *BIOMEDICAL AND HEALTH SCIENCES*

In this section, we present a small study of the citation performance of four European countries: Denmark, the Netherlands,

5 17 September 2013.

Country	1991			2008		
	FA	FA_top-1	PFA_top-1	FA	FA_top-1	PFA_top-1
Denmark	2411.5	21.0	0.9	3199.0	49.5	1.5
The Netherlands	5116.6	46.8	0.9	9055.6	122.3	1.4
Sweden	5138.0	47.2	0.9	5951.8	48.9	0.8
Switzerland	3257.2	51.0	1.6	5025.1	88.8	1.8

Table 1. *FA, FA_top-1 and PFA_top-1 for the four countries Denmark, the Netherlands, Sweden and Switzerland for the two publication years 1991 and 2008*

Sweden and Switzerland. In the study, which concerns how well represented the four countries are among the top-1 percent most cited documents in the field *Biomedical and Health Sciences*, the approach described in the preceding section is applied.

3.1 DATA AND METHODS

The delimitation of the field *Biomedical and Health Sciences* used in the study is the one applied in the CWTS Leiden Ranking 2013[6]: A (WoS) article belongs to the field *Biomedical and Health Sciences* if and only if its journal belongs to at least one the 64 different WoS journal categories listed in Table 2 (see Appendix 1). We used the query given in Appendix 1 to retrieve all documents from the WoS (provided by Thomson Reuters) belonging to *Biomedical and Health Sciences*, are of type *article*, and are published in 1991 or 2008 (databases = SCI-EXPANDED, SSCI, A&HCI). The query was performed on 1 August 2013, resulting in 634,224 retrieved articles. Let *S* be the set of these articles.

Percentile Definition 1 was used. For each of the two publication years, we retrieved the articles published in that particular year (229,369 for 1991; 404,855 for 2008). Then we applied our approach (using definition 1) to the two resulting subsets of S in order to identify the top-1 percent most cited articles. For the year 1991, the 99th percentile of the citation distribution was 255, and all articles with a citation value larger than 255, 2,284 in total, were exported to files. The 99th percentile of the citation distribution for the year 2008 was 98, leading to a target set of 4,028 exported articles. The next step was to retrieve, and export, the set of all articles in S with at least one address such that at least one of the country names *Denmark, Netherlands, Sweden* and *Switzerland* is present in the address (we used Analyze Results (Countries/Territories)). It turned out hat 50,776 articles satisfied the search condition; these were exported to files.

For each of the four countries, and for a given publication year (1991, 2008), all articles, with at least one address such that the country name is present in the address, among the top-1 percent most cited articles for the year, were identified. For each such article, the number of unique country names among its addresses was obtained, say *n*, and the country was assigned the fraction 1/*n*. Then the *number of fractionalized articles* for the country, with respect to the top-1 percent most cited articles for the year, was calculated: the sum of fractions across the top-1 percent articles for the country. Finally, the number of fractionalized articles for the country was divided by the total number of fractionalized articles for the country in the publication year (with respect to the field *Biomedical and Health Sciences*), and the result was multiplied by 100. Thus, we calculated the percent fractionalized articles for the country among the top-1 percent most cited articles for the year, relative to the total number of fractionalized articles for the

6 http://www.leidenranking.com/ranking/2013

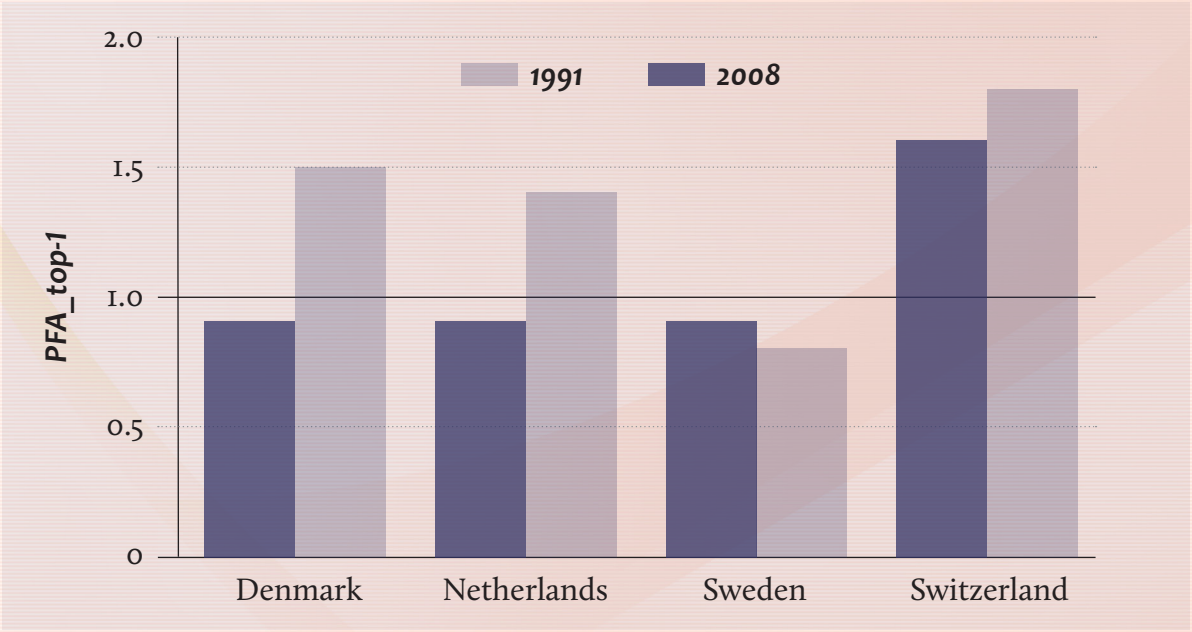


Figure 1. *PFA_top-1 for the four countries Denmark, the Netherlands, Sweden and Switzerland for the two publication years 1991 and 2008.*

country in the year. We denote this indicator by PFA_top-1. If the indicator value is greater (less) than 1, the country performs better (worse) than expected with regard to its representation among the top-1 percent most cited articles.

3.2 RESULTS

Table 1 reports, for the two considered years, the total number of fractionalized articles for the countries (denoted by *FA*), the number of fractionalized articles with respect to the top-1 percent most cited articles (denoted by *FA_top-1*) and the eight values on the indicator PFA_top-1. In Figure 1, the outcome for the latter indicator is visualized.

Denmark and the Netherlands have considerably higher values on the indicator PFA_top-1 for articles published in 2008 compared to articles published in 1991. In the latter year both countries performed worse than expected, i.e., their indicator values are less than 1. Switzerland has the highest PFA_top-1 value for both years. Sweden, which has values below 1 for both years, is the only country among the four considered that has a lower PFA_top-1 value for 2008 than for 1991. The outcome for year 2008 agrees well with the result that Sweden, across all subjects (WoS categories), though, has low proportions of highly cited documents in the publication period 2000-2011 compared to the other three countries (Karlsson and Persson, 2012).

4. CONCLUDING REMARKS

In this work, we have presented an approach for efficient online identification of the top-*k* percent most cited documents in large sets of WoS documents. We defined the top-*k* percent most cited documents in a set as those receiving strictly more than the $p^{th} = (100 - k)^{th}$ percentile of citations; four definitions of a percentile were considered. We also reported on a small study of the performance of four European countries – Denmark, the Netherlands, Sweden and Switzerland – regarding the top-1 percent most cited articles in the field *Biomedical and Health Sciences*. The study focuses on the approach in question. However, one result is that Sweden is the only country among the four considered that has a lower PFA_top-1 value for 2008 than for 1991.

On purpose we did not study the problem of determining a set of percentiles classes and attaching scores to documents

belonging to such classes, focusing instead on a simpler clearly defined problem, namely that of obtaining the top-*k* percent most cited documents in large sets of retrieved documents in one particular often-used database. One of the additional problems encountered in the general exercise is to decide how to treat uncited documents. In some cases such documents consist of more than 50% of the total set. We refer the reader to, e.g., (Bornmann 2013; Bornmann, Leydesdorff and Mutz 2013; Schreiber 2013) for a discussion of these problems.

We believe that the suggested approach can aid those colleagues that apply citation statistics, based on percentiles, and do not have access to a dedicated, bibliometric version of the WoS. By following our approach, one can efficiently identify the top-*k* percent most cited documents in a large set of retrieved WoS documents, where the retrieved documents are (maybe only partially) visible in the interface to the WoS provided by Thomson Reuters. Yet, we have to point out that when S (the set of retrieved documents) is very large, the approach does not work due to the 100,000 item interface restriction referred to in the introduction. This happens, for instance, if one is interested in the top-10 percent most cited documents in the whole database. For extreme cases like this, we have no solution because of existing database restrictions.

REFERENCES

Bornmann, L. (2013) 'How to analyze percentile citation impact data meaningfully in bibliometrics: the statistical analysis of distributions, percentile rank classes, and top-cited papers', Journal of the American Society for Information Science and Technology, 64/3: 587-95.

Bornmann, L., Leydesdorff, L., and Mutz, R. (2013) 'The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits', Journal of Informetrics, 7/1: 158-65.

Bornmann, L., Leydesdorff, L. and Wang, J. (2013) 'Which percentile-based approach should be preferred for calculating normalized citation impact values? An empirical comparison of five approaches including a newly developed citation-rank approach (P100)', Journal of Informetrics, 7/4: 933–944.

Hazen, A. (1914) 'Storage to be provided in impounding reservoirs for municipal water supply', Transactions of American Society of Civil Engineers, 77: 1539-640.

Hyndman, R. J. and Fan, Y. N. (1996) 'Sample quantiles in statistical packages', American Statistician, 50/4: 361-65.

Karlsson, S. and Persson, O. (2012) The Swedish production of highly cited papers. Stockholm: The Swedish Research Council.

Schreiber, M. (2013) 'Uncertainties and ambiguities in percentiles and how to avoid them', Journal of the American Society for Information Science and Technology, 64/3: 640-43.

The R project for statistical computing (2014), <http://www.r-project.org> accessed 18 October 2014.

Visser, M. S. and Nederhof, A. J. (2011) 'Bibliometric Analysis of Uppsala University 2007-2010'. In: Nordgren J, Andersson P, Eriksson L and Sundquist B (eds.) Quality and Renewal 2011: An overall evaluation of research at Uppsala University 2010/2011, pp. 541-67. Uppsala: Uppsala University.

Waltman, L. et al. (2011a) 'Towards a new crown indicator: an empirical analysis' Scientometrics, 87/3: 467-81.

Waltman, L. et al. (2011b) 'Towards a new crown indicator: Some theoretical considerations', Journal of Informetrics, 5/1: 37-47.

APPENDIX 1

DELIMITATION OF THE FIELD *BIOMEDICAL AND HEALTH SCIENCES* AND THE CORRESPONDING WOS QUERY

WOS CATEGORIES INVOLVED IN THE BIOMEDICAL AND HEALTH SCIENCES			
Allergy	Developmental Biology	Medicine, Legal	Primary Health Care
Anatomy & Morphology	Emergency Medicine	Medicine, Research & Experimental	Psychiatry
Andrology	Endocrinology & Metabolism	Neuroimaging	Public, Environmental & Occupational Health
Anesthesiology	Engineering, Biomedical	Neurosciences	Radiology, Nuclear Medicine & Medical Imaging
Audiology & Speech-Language Pathology	Gastroenterology & Hepatology	Nursing	Rehabilitation
Biochemical Research Methods	Genetics & Heredity	Nutrition & Dietetics	Reproductive Biology
Biochemistry & Molecular Biology	Geriatrics & Gerontology	Obstetrics & Gynecology	Respiratory System
Biophysics	Health Care Sciences & Services	Oncology	Rheumatology
Cardiac & Cardiovascular Systems	Hematology	Ophthalmology	Sport Sciences
Cell & Tissue Engineering	Immunology	Orthopedics	Substance Abuse
Cell Biology	Infectious Diseases	Otorhinolaryngology	Surgery
Chemistry, Medicinal	Integrative & Complementary Medicine	Pathology	Toxicology
Clinical Neurology	Materials Science, Biomaterials	Pediatrics	Transplantation
Critical Care Medicine	Medical Informatics	Peripheral Vascular Disease	Tropical Medicine
Dentistry/Oral Surgery & Medicine	Medical Laboratory Technology	Pharmacology & Pharmacy	Urology & Nephrology
Dermatology	Medicine, General & Intern	Physiology	Virology

Table 2. *The delimitation of the field Biomedical and Health Sciences in terms of WoS categories*

The following query was used to retrieve documents in the field Biomedical and Health Sciences, of type article, and published 1991 or 2008 (databases = SCI-EXPANDED, SSCI, A&HCI):

WC=(allergy OR anatomy & morphology OR andrology OR anesthesiology OR audiology & speech-language pathology OR biochemical research methods OR biochemistry & molecular biology OR biophysics OR cardiac & cardiovascular systems OR cell & tissue engineering OR cell biology OR chemistry, medicinal OR clinical neurology OR critical care medicine OR dentistry/oral surgery & medicine OR dermatology OR developmental biology OR emergency medicine OR endocrinology & metabolism OR engineering, biomedical OR gastroenterology & hepatology OR genetics & heredity OR geriatrics & gerontology OR health care sciences & services OR hematology OR immunology OR infectious diseases OR integrative & complementary medicine OR materials science, biomaterials OR medical informatics OR medical laboratory technology OR medicine, general & internal OR medicine, legal OR medicine, research & experimental OR neuroimaging OR neurosciences OR nursing OR nutrition & dietetics OR obstetrics & gynecology OR oncology OR ophthalmology OR orthopedics OR otorhinolaryngology OR pathology OR pediatrics OR peripheral vascular disease OR pharmacology & pharmacy OR physiology OR primary health care OR psychiatry OR public, environmental & occupational health OR radiology, nuclear medicine & medical imaging OR rehabilitation OR reproductive biology OR respiratory system OR rheumatology OR sport sciences OR substance abuse OR surgery OR toxicology OR transplantation OR tropical medicine OR urology & nephrology OR virology) AND DT=(Article) AND PY=(1991 OR 2008)

SELECTED BIBLIOGRAPHY OF THE FIRST 10 YEARS OF THE ISSI NEWSLETTER

COMPILED & INTRODUCED BY **BALÁZS SCHLEMMER**

289 titles by 104 authors from 26 countries have so far been published in the ISSI Newsletter during its 10-year-long existence. When the first issue came out, Wolfgang Glänzel editor-in-chief wrote the following vision on key objectives in his opening editorial:

"The editors sincerely hope that [...] this Newsletter will help to improve communication among the Society members by regularly informing about the Society's life and the activities and problems of their members in a rather informal way, and also help to bridge the gap between the daily flood of unfiltered information and the formal communication of scientific publishing."

It's difficult to decide whether his vision was so accurate or the newsletter's authors and editors were the ones who followed these editorial principles so precisely, but the result has proven to be very much in line with the above key objectives (see Fig 1). We have kept in-

Fig. 1 *Distribution of article types in the ISSI Newsletter (2005-2014)*

forming the members about the Society's matters (president's reports, elections, anniversary retrospections), tried to balance between the informal and formal way of communication (e.g. conference reports, interviews, introduction of awardees, scientific cartoons vs. "serious" scientific matters), and most importantly, we tried to bridge that certain gap between the entries relevant primarily because of their timeliness (news, announcements, conference calls) and those that are closer to the traditional research publishing practice and therefore provide longer-lasting scientific impact.

These certain book reviews, short communications and articles with "longer-lasting scientific impact" are the ones that have been selected for this 10th Anniversary Bibliography. The bibliography is ranked in chronological order.

Additionally, we would like to express our thanks and show our appreciation to our authors for their diligence by adding a full author index (all document types included) after the selected bibliography. The index contains co-authors as well and is ranked by author surnames.

Last but not least, and without mentioning any name, we would also like to thank the creators of all those 397 photographs, 156 charts, 96 other illustrations and 100 tables that made the Newsletter more colourful, more attractive and/or more comprehensible in the past 10 years.

Basu, A. (2005), Scientometrics / Bibliometrics in India. *ISSI Newsletter*, **1** (1), 3-3.

Braun, T. (2005), Book Review: Personalized Number Crunching in a Handbook. A Quasi Book Review. *ISSI Newsletter*, **1** (1), 5-6.

Rousseau, R. (2005), Another Look at Small Worlds. One Node Set – Two Link Structures. *ISSI Newsletter*, **1** (1), 7-7.

Leta, J. (2005), Editorial: Brazilian Science and Free Information Access. *ISSI Newsletter*, **1** (2), 1-2.

Glänzel, W., Rousseau, R. (2005), Erdős Distance and General Collaboration Distance. *ISSI Newsletter*, **1** (2), 4-5.

Glänzel, W. (2005), Book Review (Rafael Ball, Dirk Tunger: Bibliometric Analyses – Data, Facts and Methods. Basic Knowledge in Bibliometrics for Scientists, Science Managers, Research Institutions, and Universities.). *ISSI Newsletter*, **1** (2), 6-8.

Bonitz, M. (2005), The Matthew Effect for Countries (MEC): Its Impact on Information Science. *ISSI Newsletter*, **1** (3), 8-10.

Godin, B. (2005), A Centennial 1906-2006. *ISSI Newsletter*, **1** (4), 7-8.

Rousseau, R., Small, H. (2005), Escher Staircases Dwarfed. *ISSI Newsletter*, **1** (4), 8-10.

Basu, A. (2005), Book Review (Howard Bloom: Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century). *ISSI Newsletter*, **1** (4), 11-14.

Glänzel, W., Persson, O. (2005), H-Index for Price Medalists. *ISSI Newsletter*, **1** (4), 15-18.

Bar-Ilan, J. (2006), H-Index for Price Medalists Revisited. *ISSI Newsletter*, **2** (1), 3-5.

Egghe, L. (2006), An Improvement of the H-Index: the G-Index. *ISSI Newsletter*, **2** (1), 8-9.

Rousseau, R. (2006), After the Journal Impact Factor and the Web Impact Factor a Referee Factor Enters the Fray: Some Comments. *ISSI Newsletter*, **2** (2), 2-3.

Egghe, L. (2006), The New Journal of Informetrics. *ISSI Newsletter*, **2** (2), 3-4.

Rousseau, R. (2006), The World of Science: Interesting Figures. *ISSI Newsletter*, **2** (2), 4-4.

Rousseau, R. (2006), Scientific Thinking is Applied Logics, Isn't it?. *ISSI Newsletter*, **2** (2), 6-7.

Kosmulski, M. (2006), A New Hirsch-type Index Saves Time and Works Equally Well as the Original H-Index. *ISSI Newsletter*, **2** (3), 4-6.

Jang, L., Jin, B. (2006), A Co-Occurrence Study of International Universities and Institutes Leading to a New Instrument for Detecting Partners for Research Collaboration. *ISSI Newsletter*, **2** (3), 7-9.

Rousseau, R. (2006), Derek De Solla Price's Work Corrected. *ISSI Newsletter*, **2** (4), 4-4.

Rousseau, R. (2006), Preference for the Own Group Favours Largest Group Most. *ISSI Newsletter*, **2** (4), 5-5.

Glänzel, W. (2007), Editorial: Who ranks the rankers?. *ISSI Newsletter*, **3** (1), 1-2.

Lewison, G. (2007), Proposal for a Database of Biomedial Research Articles Cited by International Media Stories (BRACIMS). *ISSI Newsletter*, **3** (1), 5-5.

Jin, B. (2007), The AR-Index: Complementing the H-Index. *ISSI Newsletter*, **3** (1), 6-6.

Glänzel, W., Zimmerman, E., Bar-Ilan, J. (2007), A Scientometric Look at Scholarly Cooperation between Europe and Israel. An Explorative Study of a Changing Landscape. *ISSI Newsletter*, **3** (1), 7-10.

Thelwall, M. (2007), Web Science: What Can Information Science Contribute?. *ISSI Newsletter*, **3** (2), 21-24.

Basu, A. (2007), A Note on the Connection between the Hirsch Index and the Random Hierarchical Model. *ISSI Newsletter*, **3** (2), 24-27.

Glänzel, W. (2007), Some New Applications of the H-Index. *ISSI Newsletter*, **3** (2), 28-31.

Rousseau, R. (2007), Cha-Cha-Cha in Informetrics. *ISSI Newsletter*, **3** (3), 43-45.

Kosmulski, M. (2007), Lack of Consequence in English Translations of Universities' Names Ruins their Scientific Reputation. *ISSI Newsletter*, **3** (3), 46-48.

Guns, R. (2007), Information (Retrieval) + Metrics = Informetrics (The Google Way). *ISSI Newsletter*, **3** (4), 53-54.

Berendt, B. (2007), The Missing Link between Google Scholar and Plagiarism Prevention? How Citation Analysis Can Help Students Learn About the Nature of Knowledge. *ISSI Newsletter*, **3** (4), 55-59.

Zhou, P. (2007), The Measurement of Science and Technology in China. *ISSI Newsletter*, **3** (4), 60-63.

Rousseau, R. (2008), Triad or Tetrad: Another Representation. *ISSI Newsletter*, **4** (1), 5-7.

Zhou, P., Leydesdorff, L. (2008), China Ranks Second in Scientific Publications Since 2006. *ISSI Newsletter*, **4** (1), 7-9.

Glänzel, W. (2008), Turkey on the Way to the European Union? On a Scientific Power Rising Next Door. *ISSI Newsletter*, **4** (1), 10-17.

Wolfram, D. (2008), Is Informetrics Research on the Decline in North America?. *ISSI Newsletter*, **4** (2), 21-22.

The KNUDOP Search Group (2008), PODUNK Effect. *ISSI Newsletter*, **4** (2), 23-23.

Glänzel, W. (2008), Seven Myths in Bibliometrics About Facts and Fiction in Quantitative Science Studies. *ISSI Newsletter*, **4** (2), 24-32.

Persson, O. (2008), A Fairly Recent ACA Map of Information Science – 2003-2007. *ISSI Newsletter*, **4** (3), 37-38.

Glänzel, W., Gupta, B.M. (2008), Science in India. A Bibliometric Study of National Research Performance in 1991-2006. *ISSI Newsletter*, **4** (3), 42-48.

Rousseau, R. (2008), Luckily, Science Focuses on Achievements. Some Thoughts Related to the H-index. *ISSI Newsletter*, **4** (3), 49-50.

Schubert, A., Glänzel, W. (2008), Ternary Plots of Science in a Tripolar World. *ISSI Newsletter*, **4** (3), 51-52.

Persson, O., Schlemmer, B. (2008), Wordle: Another Visualisation. *ISSI Newsletter*, **4** (3), 53-53.

Persson, O. (2008), The Role of Classics in Social Science. *ISSI Newsletter*, **4** (4), 63-63.

Glänzel, W. (2008), What are Your Best Papers?. *ISSI Newsletter*, **4** (4), 64-67.

The Stimulate 8 Group (2009), The Web of Knowledge Versus Google Scholar: Presence of Non-western Countries. *ISSI Newsletter*, **5** (1), 3-6.

Zhiping, Y., Shu, F. (2009), Research on Patent-bibliometric Analysis in NSLC (National Science Library of Chinese Academy of Sciences). *ISSI Newsletter*, **5** (1), 7-9.

Das, A.K. (2009), Open Access to Research Literature in India: Contemporary Scenario. *ISSI Newsletter*, **5** (1), 9-14.

Liu, Y. (2009), Peripheral Articles. *ISSI Newsletter*, **5** (1), 15-16.

Bhattacharya, S. (2009), Book Review (Caroline S. Wagner: The New Invisible College Science for Development). *ISSI Newsletter*, **5** (2), 28-30.

Bar-Ilan, J. (2009), Informetrics in the Eyes of Web. *ISSI Newsletter*, **5** (2), 30-34.

Persson, O. (2009), Is High Impact Research Domestic or International?. *ISSI Newsletter*, **5** (2), 35-35.

Braun, T. (2009), The Ranking Game on the Top the Competitive World of Nations in Scientific Research. *ISSI Newsletter*, **5** (2), 36-40.

Waltman, L., van Eck, N.J. (2009), A Simple Alternative to the H-Index. *ISSI Newsletter*, **5** (3), 46-48.

Schubert, A. (2009), Multi- and Interdisciplinarity in Medical and Veterinary Literature: Approaches and Assertions. *ISSI Newsletter*, **5** (3), 49-52.

Schlemmer, B., Glänzel, W. (2009), Science in a Changing Europe: East vs. West. National Scientific Profiles By Subject Fields. *ISSI Newsletter*, **5** (3), 52-58.

Bolaños-Pizarro, M., Thijs, B., Glänzel, W. (2009), A Comparative Scientometric Study of Cardiovascular Research in Spain. *ISSI Newsletter*, **5** (4), 70-78.

Guns, R., Rousseau, R. (2010), New Journal Impact Indicators Take References into Account: A Comparison. *ISSI Newsletter*, **6** (1), 9-14.

Chand, P. (2010), Indian Citation Index (ICI): Method, Material and Construction Approach. *ISSI Newsletter*, **6** (1), 15-21.

Sivertsen, G. (2010), A Performance Indicator Based on Complete Data for the Scientific Publication Output at Research Institutions. *ISSI Newsletter*, **6** (1), 22-28.

Bar-Ilan, J. (2010), A Follow-up on the H-Index of Price Medalists. *ISSI Newsletter*, **6** (2), 39-43.

Glänzel, W. (2010), What is the Impact of Fraudulent Literature?. *ISSI Newsletter*, **6** (2), 44-47.

Labbé, C. (2010), Ike Antkare, One of the Great Stars in the Scientific Firmament. *ISSI Newsletter*, **6** (2), 48-52.

Rousseau, R. (2010), Ci-Speed: A Hirsch-type Citation Speed Index for a Set of Articles. *ISSI Newsletter*, **6** (3), 62-65.

Glänzel, W., Zhang, L. (2010), A Demographic Look at Scientometric Characteristics of a Scientist's Career. *ISSI Newsletter*, **6** (3), 66-84.

Persson, O. (2011), Price Award Mapping. *ISSI Newsletter*, **7** (1), 4-4.

Rousseau, R., Ye, F.Y. (2011), Subgraphs Derived from the Hirsch Core in Undirected, Unweighted Networks. *ISSI Newsletter*, **7** (1), 5-9.

Rousseau, R., Leydesdorff, L. (2011), Simple Arithmetic Versus Intuitive Understanding: The Case of the Impact Factor. *ISSI Newsletter*, **7** (1), 10-14.

Srivastava, D., Kant, R. (2011), Directory and Database of Indian Science, Technology and Medical Periodicals (DIP). *ISSI Newsletter*, **7** (1), 15-19.

Glänzel, W. (2011), How to Obtain Erdős Number 1 in the 21th Century? Some Thoughts on Posthumous Co-authorship. *ISSI Newsletter*, **7** (1), 20-22.

Prathap, G. (2011), A Tale of Two Cities: A Scientometric Comparison of Beijing and New Delhi Using an iCX Map Representation. *ISSI Newsletter*, **7** (2), 28-30.

Zhang, L., Glänzel, W. (2011), On the Role of Proceedings Papers Published in Journals. *ISSI Newsletter*, **7** (2), 31-36.

van Eck, N.J., Waltman, L. (2011), Text Mining and Visualization Using VOSviewer. *ISSI Newsletter*, **7** (3), 50-54.

Egghe, L. (2011), Benford's Law is a Simple Consequence of Zipf's Law. *ISSI Newsletter*, **7** (3), 55-56.

Levitt, J.M. (2011), Preliminary Findings on whether it is Good Value for Money to Fund Larger Research Groups. *ISSI Newsletter*, **7** (3), 57-62.

Mahbuba, D., Rousseau, R. (2011), Bangladesh' Publication Barycentre. *ISSI Newsletter*, **7** (3), 63-68.

Czerwon, H-J. (2011), Jan Vlachý's Scientific Estate at the K.U. Leuven. *ISSI Newsletter*, **7** (4), 83-84.

Glänzel, W., Thijs, B. (2011), Research in Emerging Fields: Who takes the lead?. *ISSI Newsletter*, **7** (4), 85-95.

Glänzel, W., Heeffer, S. (2012), Editorial: The Seven-Year Itch. *ISSI Newsletter*, **8** (1), 1-3.

Persson, O. (2012), A New Measure of Globalization – The Co-citation Triangle Perimeter (CTP). *ISSI Newsletter*, **8** (2), 20-21.

Ye, F.Y. (2012), H-inconsistency Is Not an Issue in Dynamical Systems. *ISSI Newsletter*, **8** (2), 22-24.

Rousseau, R. (2012), Into the Future. *ISSI Newsletter*, **8** (2), 25-28.

Glänzel, W., Thijs, B., Zhang, L. (2012), Is Scientometrics a 'Metrics' Indeed?. *ISSI Newsletter*, **8** (2), 29-33.

Leta, J. (2012), Brazilian Scientometrics: From Little to Big?. *ISSI Newsletter*, **8** (3), 34-37.

Rousseau, R., Hu, X. (2012), Predicting the h-index. *ISSI Newsletter*, **8** (3), 53-55.

Schubert, A. (2012), Dear Sir. *ISSI Newsletter*, **8** (3), 56-57.

Rousseau, R. (2012), Thoughts about the Activity Index and its Formal Analogues. *ISSI Newsletter*, **8** (4), 73-75.

Persson, O. (2013), H-Index on Everything. *ISSI Newsletter*, **9** (1), 5-5.

Mêgnigbêto, E. (2013), Scientific Research in West Africa: A Global View (2001-2010). *ISSI Newsletter*, **9** (1), 6-11.

Woeginger, G.J. (2013), How Useful Is the E-Index?. *ISSI Newsletter*, **9** (1), 12-15.

Heeffer, S., Thijs, B., Glänzel, W. (2013), Are Registered Authors More Productive?. *ISSI Newsletter*, **9** (2), 29-32.

Rousseau, R. (2013), The Standard Synchronous Impact Factor versus the Diachronous One. *ISSI Newsletter*, **9** (2), 33-36.

Wouters, P., Glänzel, W., Gläser, J., Rafols, I. (2013), The Dilemmas of Performance Indicators of Individual Researchers – An Urgent Debate in Bibliometrics. *ISSI Newsletter*, **9** (3), 48-53.

Schubert, A. (2013), A Follow-up Study of Title Word Bisociations in Inorganica Chimica Acta. *ISSI Newsletter*, **9** (3), 54-55.

Lietz, H., Rousseau, R. (2013), A Mathematical Relationship between Growth and Scaling in an Exponentially Growing Bibliographic System. *ISSI Newsletter*, **9** (4), 74-76.

Persson, O. (2014), The Author Affiliation Syndrome—Does It Really Matter?. *ISSI Newsletter*, **10** (1), 12-13.

Mêgnigbêto, E. (2014), Information Flow between West African Triple Helix Actors. *ISSI Newsletter*, **10** (1), 14-20.

Persson, O., Glänzel, W. (2014), Growing h-index for Some Price Medalists. *ISSI Newsletter*, **10** (1), 21-23.

Levitt, J.M. (2014), Nobel Prize Winner Publishes in ISSI Proceedings. *ISSI Newsletter*, **10** (1), 24-26.

Schubert, A., Glänzel, W. (2014), Editorial: Plagiarism: A Non-Apology. *ISSI Newsletter*, **10** (2), 27-28.

Rousseau, R. (2014), Advanced Search in Thomson Reuters' Web of Science. *ISSI Newsletter*, **10** (2), 43-45.

Rafols, I., de Rijcke, S., Wouters, P. (2014), The Leiden Manifesto in the Making. Full Report of the Plenary Session at the 2014 STI Conference in Leiden on Quality Standards for Evaluation: Any Chance of a Dream Come True?. *ISSI Newsletter*, **10** (3), 60-62.

Rousseau, R., Zhang, L. (2014), How to Determine the H-index of a Set of Publications in the WoS?. *ISSI Newsletter*, **10** (3), 63-65.

Ahlgren, P., Persson, O., Rousseau, R. (2014), An Approach for Efficient Online Identification of the Top-k Percent Most Cited Documents in Large Sets of Web of Science Documents. ISSI Newsletter, **10** (4), 81-89.

AUTHOR INDEX OF THE FIRST 10 YEARS OF THE ISSI NEWSLETTER