



ISSI NEWSLETTER

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

#38 / VOLUME 10 NUMBER 2
JUNE 2014

CONTENTS

EDITORIAL

A. Schubert
& W. Glänzel:
Plagiarism:
A Non-Apology
page 27

CONFERENCE

Invitation to the
19th Nordic Workshop
on Bibliometrics and
Research Policy
page 29

Call for Papers:
Scientometrics2014
International Confer-
ence & Trade Fair
page 31

International Semi-
nar on the Quantita-
tive and Qualitative
Study of Science
and Technology
“Professor Gilberto
Sotolongo Aguilar”
– A conference report
page 33

REPM
– A conference report
page 38

ARTICLES

R. Rousseau:
Advanced Search in
Thomson Reuters’
Web of Science
page 43

EDITORIAL

PLAGIARISM: A NON-APOLOGY

If somebody spends several decades in the world of scientific publica-
tions as author, referee and editor, then cannot escape encountering
certain ethical issues.

Researchers are fallible
humans, after all, not
free of vices like van-
ity, envy or jealousy.
The range is wide from
slightly tendentious se-
lection among data to
rudely falsified results,
from dextrous overlooking
relevant works of the con-
current group to straight
plagiarism.

The phenomenon is
far not new, it can cer-
tainly be tracked back
to the ancient Greeks or
even farther. As contem-
porary developments, two



ANDRÁS
SCHUBERT^a



WOLFGANG
GLÄNZEL^{a,b}

^a Library of the Hungarian Academy
of Sciences, Dept. Science Policy &
Scientometrics, Budapest, Hungary

^b K.U. Leuven, Centre for R&D Monitor-
ing and Dept. MSI, Leuven, Belgium

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

Accepted contributions are moderated by the board. Guidelines for contributors can be found at <http://www.issi-society.info/editorial.html>.
Opinions expressed by contributors to the Newsletter do not necessarily reflect the official position of ISSI. Although all published material is expected to conform to ethical standards, no responsibility is assumed by ISSI and the Editorial Board for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material therein.



factors can be highlighted that apparently act in opposing directions.

One is the umpteenth times condemned “publish or perish” pressure.

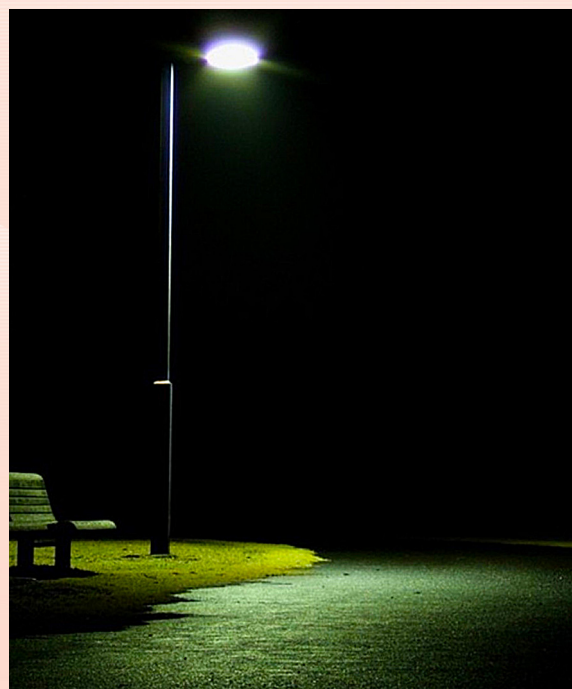
Pressure is on the authors to “produce” results by hook or by crook, and this may effectively lower their moral barriers. Pressure is on the referees to keep tight deadlines to give explicit and unambiguous advices even at the cost of less immersion in the topic. And pressure is on the editors to fill the prescribed volume of the journal with articles that will presumably be requested, downloaded, sold and cited in accordance with the Publisher’s objectives. This triple pressure obviously increases the chances of the publication of papers of dubious ethical standing.

On the other hand, there is an ever evolving toolkit helping to detect various forms of deviant publication behavior. Text analysis software is able to pinpoint duplicate publications (even in different languages), or similarities in the text of articles of the same or different authors. In the latter case, if the earlier text is not properly referenced, the suspicion of plagiarism can justly be raised. A higher percentage degree of textual similarity may, however, stem from various innocent causes: the abundant occurrence of standard definitions, idiosyncratic methodological descriptions, shared references, etc. The suspicion, therefore, must not lead to automatic verdict. Like bibliometrics for research evaluation, the quantitative measures may help but not substitute careful human deliberation.

We hasten to assert: any form of plagiarism is an ethical misdemeanor and is unacceptable. But, as it is justly stressed in a recent Nature note [1] and in the subsequent comments, the sanctions should be nuanced, and the wrongdoing should be placed in proper framework. It seems unjust to name plagiarism the main evil just because its possible presence can so easily be found even by lay outsiders. Each case must be, of course, carefully investigated by a proper expert body whose expertise is,

at the same time, painfully needed to discern other, sometimes more severe, ethical issues that not so readily reveal themselves.

It may sound absurd, but plagiarism, however severe offence is against those deserving credit, appears not to harm the body of scientific knowledge itself. A copy-pasted mathematical theorem remains true, a chemical synthesis method will work even if it was found in an expropriated document. On the other hand, manipulated results, fraudulent evidences, inadequate statistical inferences may cause serious harm if incorporated into the body of knowledge. We can only wish if we would have half as effective tools to unveil them as we have for plagiarism.



Until then, we are like the guy who tries to find his lost key near to the lamp post. Not as if he would have lost it there, but because there is light.

REFERENCE

- [1] Chaddah, P., Not all plagiarism requires a retraction. *Nature* 511, 127 (07 July 2014), doi:10.1038/511127a;
<http://www.nature.com/news/not-all-plagiarism-requires-a-retraction-1.15517>

INVITATION TO THE 19TH NORDIC WORKSHOP ON BIBLIOMETRICS AND RESEARCH POLICY

25—26 SEPTEMBER 2014
REYKJAVÍK, ICELAND

The 19th Nordic Workshop on Bibliometrics and Research Policy will be held in Reykjavík on September 25th and 26th 2014. Venue is at the campus of University of Iceland.

Number of researchers on bibliometrics and research policy have submitted very interesting abstracts to be further addressed at the workshop.



Reykjavík with the Sun Voyager. Photo courtesy of © Balázs Schlemmer | s-press.hu

Unfortunately it is not possible for us to obtain any more abstracts. However the Workshop is open for everyone and our web site <http://www.rannis.is/bibliometrics> and the information in menu on the right hand side are useable for all interested.

Workshop participants are invited to a welcoming reception on the date of their arrival, Wednesday 24 September at Hotel Reykjavik Centrum, located in Aðalstræti 16 in the old city centre. Participants can register for the workshop at the reception.

On Thursday 25 September, the workshop starts at 9:00 and ends at 16:00. At 16:30 we foresee presentations by supplier of bibliometric data until 17:30.

KEYNOTE SPEAKER:

Dr. Sybille Hinze, Institute for Research Information and Quality Assurance, Berlin

COSTS OF PARTICIPATION, TRAVEL AND ACCOMMODATION

Participation to the workshop is free. Travel and accommodation have to be arranged and sponsored by the participants themselves.

CONTACT

For further information, please contact the coordinator of the workshop:

Thorvald Finnbjörnsson: thorvald@rannis.is

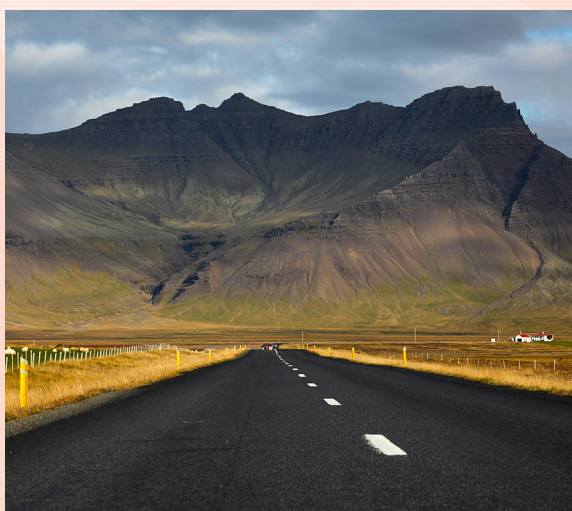
REGISTRATION TO THE WORKSHOP

Please register at this website:

<http://www.rannis.is/bibliometrics/>

Please find registration in menu.

Information on the workshop, hotel, workshop programme, accommodation, venue etc. have been posted there.



Social program: Guided visit to the Reykjanes peninsula. Photos: courtesy of © Balázs Schlemmer | s-press.hu

SCIENTOMETRICS2014

INTERNATIONAL CONFERENCE & TRADE FAIR

UNIVERSITY LIBRARY REGENSBURG, GERMANY
SEPTEMBER 24—26, 2014

CALL FOR PAPERS

ALTMETRICS: WHAT COMES AFTER CLASSICAL BIBLIOMETRICS?

The university library Regensburg is hosting the 2nd International Scientometrics Conference and Trade Fair “Scientometrics2014” from September 24th to 26th, 2014. The conference focuses on the relationship between classical scientometrics and new forms such as altmetrics.

Web2.0 is changing academic communication. Can scientometrics keep pace? Where and how can scientometric procedures be utilized in future? How will additional procedures regarding altmetrics be developed and applied in order to illustrate this transition in the age of digital change? This exciting and innovative subject will be discussed during the conference.



Universität Regensburg
UNIVERSITÄTSBIBLIOTHEK



The conference is addressed to bibliometricians, librarians, scientists from all fields, information providers and decision makers in science and research.



Regensburg riverbank panorama at sunset. Photo: © Karsten Dörre (grizurgbg) | Wikipedia Creative Commons

The presentations and conference workshops will be held in both German & English.

CALL FOR PAPERS AND POSTERS

We invite you to submit articles and posters on the following main subjects:

- ▶ New Methods
- ▶ Altmetrics
- ▶ Webometrics
- ▶ Libmetrics
- ▶ Scientometrics / Altmetrics for Research Management
- ▶ Scientometrics / Altmetrics in Research Funding
- ▶ Open Science: Experiences and Perspectives
- ▶ Text Mining
- ▶ Metrics for Interdisciplinarity
- ▶ Information Management and Scientometrics
- ▶ Science Policy
- ▶ Data Pool for Scientometrics / Altmetrics
- ▶ Quality Indicators
- ▶ Patent Analyses

DEADLINE & MORE INFO

All submissions for Bibliometrics2014 are expected to showcase original research findings from a scientific or economic perspective. The submitted abstracts for presentations will be examined by the program's committee.

The fully formulated submissions will be electronically published in the Open Access Magazine *Bibliometrie – Praxis und Forschung* www.bibliometrie-pf.de by the University Library in a proceedings volume.

Notification of Acceptance:
August 15, 2014

Submission of the Fully Formulated Article: September 1, 2014; Early submissions

will be immediately incorporated into the electronic proceedings volume.

SOCIAL PROGRAM

Reception in the Gothic Imperial Chamber

The town of Regensburg welcomes all visitors of Scientometrics2012 in the Gothic Imperial Chamber located in the Old Town Hall of Regensburg. This important medieval chamber shows magnificent mural painting of the 16th century as well as the imperial throne.

Conference Dinner at restaurant House Heuport

All visitors of Scientometrics2012 are invited to take part in our conference dinner held on the exceptional premises of the House Heuport in Regensburg. The cost of the conference menu will be included in the conference fee. Drinks will be billed separately.

Guided tour through the Botanical Garden of Regensburg University

Come and watch more than 5.500 plant species from all over the world in our large Botanical Garden. Plants usually growing in various climate zones and different continents are inviting you for a relaxing walk.

Guided City Tour through Regensburg

Patricians and bishops, demons and saints, citizens and craftsmen have left their mark on Regensburg's colourful history. Discover the major landmarks and monuments of one of the most important towns during the Middle Ages and listen to vivid stories and legends of the people of Regensburg.

INFORMATION & REGISTRATION

Web: www.bibliometrie2014.de

Host: Universitätsbibliothek Regensburg
Universitätsstr. 31,
93053 Regensburg,
Germany

[bibliometrie\(at\)bibliothek.uni-regensburg.de](mailto:bibliometrie(at)bibliothek.uni-regensburg.de)

INTERNATIONAL SEMINAR ON THE QUANTITATIVE AND QUALITATIVE STUDY OF SCIENCE AND TECHNOLOGY “PROFESSOR GILBERTO SOTOLONGO AGUILAR”

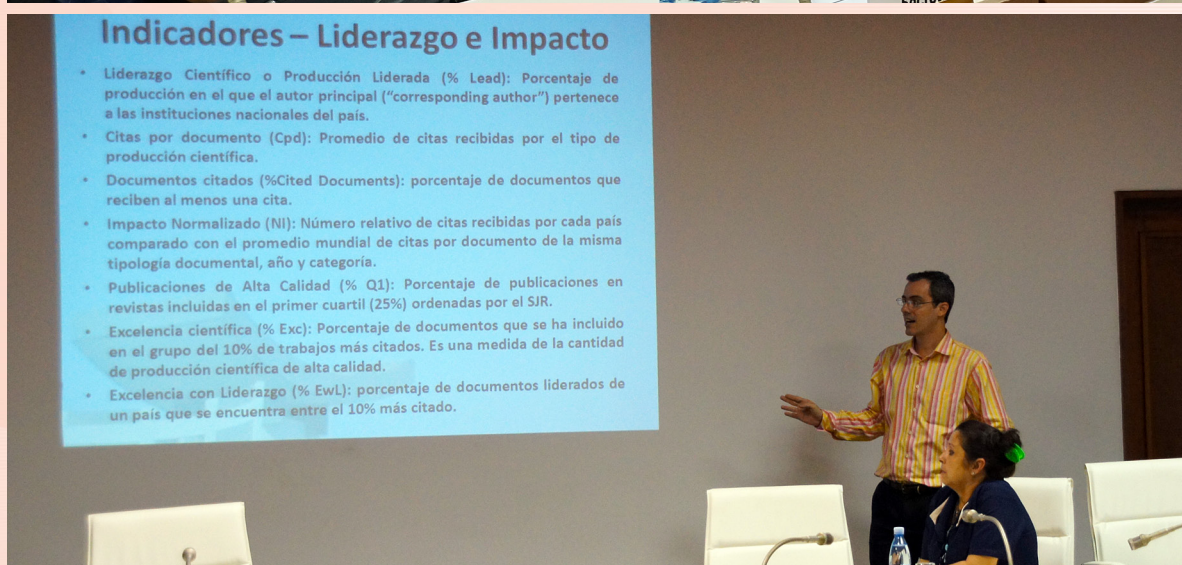
– A CONFERENCE REPORT

As happens every two years in the month of April, the *International Seminar on the Quantitative and Qualitative Study of Science and Technology “Professor Gilberto Sotolongo Aguilar”* now in its seventh edition, was held at the Palacio de Convenciones in Havana, Cuba. The Seminar began on April 15th with

the inauguration of the poster session and concluded on April 17th, the programme included 47 presentations, 28 oral and 19 posters. As in previous years, there was an important number of studies from Cuba and Mexico but other countries were also represented such as Belgium-Hungary, Brazil,

Photo courtesy of © Balázs Schlemmer | spress.hu





Peru and Spain. The great majority of presenters were once more from universities.

The keynote speaker on Wednesday the 16th was again Dr. Wolfgang Glänzel with a

talk entitled “Analysis of co-authorship patterns at the individual level”. He explained how publication activity, citation impact and communication patterns, in general,



change in the course of a scientist's career. Mobility and radical changes in a scientists' research environment or profile are among the most spectacular factors that affect individual collaboration patterns. Although bibliometrics at this level should be applied with the utmost care, characteristic patterns of research collaboration of individual scientists and its change in the course of the career can be well depicted using bibliometric methods. A bundle of indicators and network tools are chosen to follow up the evolution and to visualize and quantify collaboration and performance profiles of individual researchers. These methods are, however, designed to supplement expert-opinion based on other qualitative assessment and should not be used as stand-alone evaluation tools.

The inaugural conference for the session on Thursday the 17th was presented by Dr. Rogério Mugnaini from Brazil, the title of which was "Scientific communication in Brazil (1998—2012): national infrastructure and internationalization". The speaker addressed the reasons why Brazilian scientific production is growing in the international scenario. The study grouped together production from the Web of Science (WoS) and SciELO databases, eliminating duplicates and considering different areas and the recent internationalization of Brazilian science and journals. The journal flow between the SciELO and WoS databases showed that the number of journals exclusively indexed in SciELO increased in the period, except in the fourth triennium (2007—2009) when several journals were also indexed in the



WoS. More than 60% of national production was published in international journals (WoS). Moreover, the proportion of papers published in national journals in the five triennia, is increasing in areas where production is mainly in international journals, such as Biological, Earth and Exact Sciences; and decreasing in areas traditionally publishing in national journals, such as Health, Social, Human and Agricultural Sciences.

Discussions of all presentations were lively and enriching with speakers and audience exchanging benchmarks, ideas, comments and recommendations. The poster session was well-attended where discussions centred on metric indicators as tools to evaluate and elevate the quality of scientific publications.

Presentations covered diverse themes, for example university rankings, scientific communication studies, public health, marketing in information science, astronomy, neurosciences, among others. Of special interest were those related to issues of particular concern to the community of specialists from the Latin American region such as the quality of nationally generated databases and the representation of national publications in the most well-known databases such as WoS and Scopus. Both aspects were considered significant to the use of mainstream sources for measuring local science. Another widely-commented topic

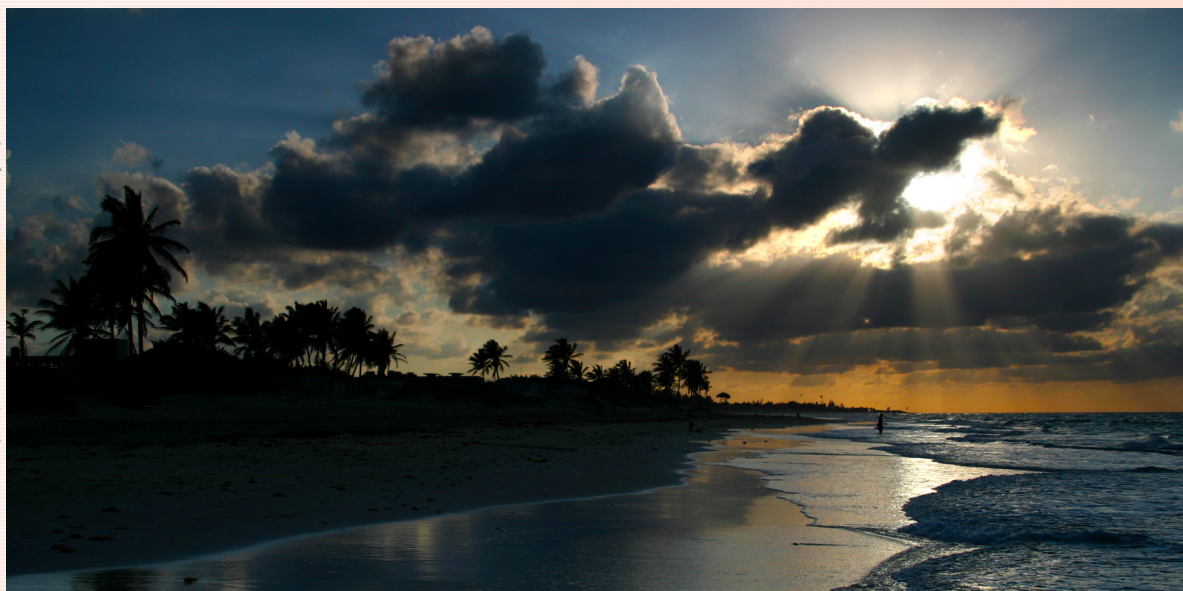
was the decrease in the scientific and technological production of several of the countries in the region in the last three years.

The importance of continuing to hold the Seminar was agreed upon as well as the need to following up on the Network of Professionals in Metric Studies. Consensus was reached on the following actions:

- ▶ Incorporation of a forum or discussion list on scientific publications from a metric standpoint to include topics such as quality indicators, local databases, visibility, etc. Scientific publications were a major concern in several of the sessions.
- ▶ Organization of a workshop for the presentation and demo of software for information analysis, these could be commercial, open source or locally produced.
- ▶ Creation of a space for the Seminar in Facebook to encourage the integration of new colleagues interested in these topics.

An important outcome of the VII Seminar will be the edition of a monographic issue of the Brazilian journal *Transformação* containing a selection of studies presented during the event. *Transformação* is published

Photo courtesy of © Balázs Schlemmer | spress.hu



by the Faculty of Library Science and the Centre for Applied Social and Human Sciences of the Pontifical Catholic University of Campinas. The journal is indexed in the WoS, Scopus and SciELO, among others.

The organizing committee will be grateful to receive suggestions and comments on this brief review of the VII Seminar and hopes to count on the participation of more countries and the inclusion of a wider range of topics in future editions.

ORGANIZING COMMITTEE

- ▶ Jane M. Russell
Universidad Nacional Autónoma de México, México
jrussell@unam.mx
- ▶ Maria Victoria Guzmán
Instituto Finlay, Cuba
mvguzman@finlay.edu.cu
- ▶ Isidro Aguillo
Consejo Superior de Investigaciones Científicas. España
isidro.aguillo@cchs.csic.es
- ▶ Francisco Collazo Reyes
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México
fcollazo@fis.cinvestav.mx
- ▶ Rogério Mugnaini
Universidade de São Paulo, Brasil
rogerio.mugnaini@gmail.com

Photo courtesy of © Balázs Schlemmer | spress.hu





REPM – A CONFERENCE REPORT

CONFERENCE CONTEXT

In later 2010 the Carnegie Corporation of New York (CCNY) awarded a grant under the title of “Research Management Tool in Support of Academic Research in South Africa, Ghana and Uganda”. Essentially the grant enabled four universities, namely Makerere University (MU, Uganda) and the Universities of Cape Town (UCT, South Africa), Ghana (UG, Ghana), Pretoria (UP, South Africa) and Witwatersrand (Wits, South Africa), to purchase licences to research evaluation tools and thereby make a contribution to their research development. Two tools were purchased: UM, UP and Wits purchased Thomson Reuters’ InCites and UCT and UG purchased Elsevier’s SciVal.

One of the conditions of the grant was to organise and run a collective workshop



developing skills related scientometrics. Scientometrics is the study of measuring and analysing research outputs and uses bibliometrics, the measurement of the impact of research publications. This need was implemented through the organisa-

tion of the Research Evaluation and Performance Measurement Conference, REPM 2014, which was hosted by UCT in Cape Town between 5 and 7 May 2014.

The conference brought together interested parties including universities academ-



ics, research administrators and managers, and librarians and data administrators. The programme focused on current and new approaches to evaluate university research performance. The conference concluded with a workshop on the use of scientometrics in research evaluation led by Professor Johann Mouton of the Centre for Research and Evaluation, Science and Technology (CREST), Stellenbosch University.

THE SCIENTOMETRIC TOOLS

The following descriptions of the tools purchased were taken from the companies webpages.

- “*InCites*™ is a customized, web-based research evaluation tool that allows you to analyze institutional productivity and benchmark your output against peers worldwide. **With customized citation data, global metrics, and multidimensional profiles** on the leading research institutions, *InCites* gives comprehensive insight into your institution’s performance. And with robust

visualization and reporting tools, you can create and share reports quickly and easily”. <http://researchanalytics.thomsonreuters.com/incites/>

- “Elsevier’s new generation of SciVal offers quick, easy access to the research performance of 4,600 research institutions and 220 countries worldwide. A ready-to-use solution with unparalleled power and flexibility, SciVal enables you to navigate the world of research and devise an optimal plan to drive and analyze your performance. SciVal builds on Elsevier’s extensive experience over several years working with many leading institutions worldwide via SciVal Spotlight and SciVal Strata”. <http://www.elsevier.com/online-tools/research-intelligence/products-and-services/scival>

DESCRIPTION OF CONFERENCE

The conference was held in the Doubletree by Hilton Upper Eastside Hotel, Cape Town, South Africa and was attended by

121 delegates. Of these 99 delegates came from South Africa, 8 from other African Countries, 11 from Europe, 1 from the UAE and 2 from Australia.

The conference programme was focused on the following themes:

1. Research measurement practice (tools, methods and approaches in support of research evaluation and measurement)
2. Collaboration (between libraries, research offices and other university support services)
3. Research support (by libraries and other university entities)
4. Institutional repositories (how they are supporting dissemination of research output, could also include policies, platforms, and possibilities of linking these to research data management initiatives)
5. Research management tools (what tools are available and how they are being used in research performance management and evaluation)
6. Case studies – each of the five partners participating in the CCNY funded project to enhance research in their institutions made joint presentations (Research Office and Library).

All presentations can be viewed at: <http://www.repm2014.co.za/index.php/2014-presentations>

INVITED SPEAKERS

Ms Robyn Drummond is the Manager of the Science, Engineering and Medicine Unit at University of New South Wales (UNSW) Library. This unit delivers information services to academics, researchers and students within those three large faculties at UNSW, Sydney, Australia.

Prof. Dr. Stefan Kuhlmann, University of Twente, Chair of the Department of Science Technology and Policy Studies, is a teacher, researcher, and intellectual in the field of science, technology and innovation governance studies. Until 2006 he was

managing director of the Fraunhofer Institute for Systems Innovation Research (ISI), Germany, and Professor of Innovation Policy Analysis at the Copernicus Institute, University of Utrecht, The Netherlands.

Dr A van Jaarsveld is Chief Executive Officer of the National Research Foundation. His career in research, teaching and leadership include academic and management positions at the Universities of Pretoria and Stellenbosch, as Dean of Science, Adjunct Professor: Environmental Studies Programme at Dartmouth College, USA, and CEO of the National Research Foundation.

Dr Eme Owoaje is a public health physician and is currently the Director of the Research Management Office at the University of Ibadan. She has participated in a number of research administrative capacity building courses and study visits to African and US institutions. Dr Owoaje is the leader of the Development of Research Uptake in Sub-Saharan Africa Project at the University of Ibadan.

OUTCOMES OF CONFERENCE

From a research management point of view the following are some of the important learning points that were raised by the conference discussions.

- ▶ Scientometric data need to be used with great care. If used without a thorough understanding of source or meaning it can lead to false comparisons and therefore poor decision making.
- ▶ Scientometric data are useful inputs in the writing of annual research reports.
- ▶ Scientometric data reflecting the effectiveness academic collaborations is a useful input into the strategic selection of future research partners.
- ▶ Bibliometric data can be and are used as an input into funding decisions, however peer review is not likely to be replaced anytime soon by these data.
- ▶ Bibliometric data gives a picture of past performance but points to ways of increasing impact of future research.



- ▶ Bibliometric data are invaluable in the identification of an institutions research strengths and weaknesses.
- ▶ Scientometrics is spreading its area of influence beyond rigorous bibliometrics to include alternative measures e.g. number of down loads, number of site hits. These are so called 'altmetrics'. Even these less rigorous data are useful in creating a picture of research impact.
- ▶ A close and supportive collaboration between the university research office and the university library can be a powerful force in increasing research output, visibility and impact. It is paramount for all research development strategies.

CONCLUSION

The 2014 REPM Conference was the first of its kind on the African continent to bring together a wide range of stakeholders involved

in institutional research performance management and evaluation. As noted by the REPM Project Director, Felix Ubogu, "The overwhelming response to the conference confirms the need to engage in discussions and share experiences on ways that research performance measurement and evaluation can be used to assess as well as improve research productivity by researchers, institutions or a country". It is hoped that the interest and enthusiasm evident at the conference will continue to gain momentum on the continent especially as global rankings become important benchmarks for universities and other educational institutions.

ACKNOWLEDGEMENTS

The generous grant from the Carnegie Corporation of New York made it possible to present the 2014 REPM Conference and provide the 5 partner institutions with essential modern web-based research management tools.

ADVANCED SEARCH IN THOMSON REUTERS' WEB OF SCIENCE



RONALD ROUSSEAU

University of Antwerp IOIW-IBW
KU Leuven, Dept. of Mathematics
ronald.rousseau@uantwerpen.be
ronald.rousseau@kuleuven.be

Abstract: A short overview of search tools in Advanced Search is provided, with special attention to lemmatization.

INTRODUCTION

Documentary searches in Thomson Reuters' Web of Science are best performed using the Advanced Search functionality. Within this frame one can adapt the language of publication and the type of document (filtering). A set of tags are available as prefixes (see Table 1). As these are not the topic of this contribution I refer the interested reader to Thomson Reuters' documentation.

Moreover, the following Boolean operators can be used: AND, OR, and NOT. The AND-operator is used when it is required that both (or more) terms be present in retrieved records; the OR-operator is used when at least one term from the OR-search

string should be present; the NOT-operator is used when it is required that all records containing the term after NOT are excluded from the search.

SAME and NEAR/x are proximity operators. SAME is used within the address field, requiring that search terms refer to the same address. NEAR/x is used to find records where the terms joined by the operator are within a specified number of words (indicated by the x in /x) of each other. This is true even when the words are across different fields. The symbol x specifies the maximum number of words that separate the terms. Without this specification the system finds records where the terms joined by NEAR are within 15 words of each other.

TS = Topic	TI = Title	AU = Author
AI = Author Identifiers	GP = Group Author	ED = Editor
SO = Publication Name	DO = DOI	PY = Year Published
CF = Conference	AD = Address	OG = Organization—Enhanced
OO = Organization	SG = Suborganization	SA = Street Address
CI = City	PS = Province/State	CU = Country
ZP = Zip/Postal Code	FO = Funding Agency	FG = Grant Number
FT = Funding Text	SU = Research Area	WC = Web of Science Category
IS = ISSN/ISBN	UT = Accession Number	

Table 1. Tags available when using the Advanced Search functionality

MORE FUNCTIONALITY

Besides the search options clearly shown on the Advanced Search webpage, more options are available by truncation and the use of wildcards. This gives more control in the retrieval of plurals and variant spellings.

- ▶ `term*` results in the retrieval of records in which this term occurs followed by zero to many characters;
- ▶ `term?` results in the retrieval of records in which this term is followed by exactly one character;
- ▶ `term??` results in the retrieval of records in which this term is followed by exactly two characters;
- ▶ `term$` results in the retrieval of records in which this term is followed by zero or one character(s);
- ▶ The symbols `?` and `$` can also be used as wildcards inside a search term: using `t??th` in a search results in records including the terms *teeth* or *tooth*, but also *truth* (for instance).

SPELLING VARIANTS AND LEMMATIZATION

Now we come to the main purpose of this note. In the Quick Reference Guide for the Web of Science (`webofscience_qrc_en.pfd`) we find the following:

“British/English spellings are searched automatically.”

Under the heading lemmatization it is explained that

“Lemmatization automatically helps find variations by stemming for plurals (even complex plurals like tooth/teeth) and searching different verb tenses (run/running) and degrees of comparison (big finds bigger and biggest). Lemmatization can be turned off by enclosing terms in quotation marks.”

Let us try this by performing a search; we note that the numbers of retrieved records depend on the version of the WoS which is locally available and on the date of the search (here May 22, 2014). They are just given to illustrate the retrieval process and refer to the WoS as available in Flanders at that moment.

First, we do a check to see if indeed British and American spelling are searched for automatically. As an example we consider:

`TS=harbor` yields 80,894 results (Set #1).

The search `TS=harbour` also yields 80,894 results (Set #2)

Clearly the WoS automatically unifies English and American spelling. Of course, one example does not prove anything, but at least there is no reason to doubt that retrieval works as announced by Thomson Reuters.

Next we check the behaviour of lemmatization.

`TS="harbor"` yields 23,960 results (Set #3) and

`TS=("harbor" OR "harbour")` yields 33,665 results (Set #4).

Trying to find the content of Set #1 we continue:

TS=("harbor" OR "harbour" OR "harbors" OR "harbours") yields 43,297 results (Set #5) and TS=("harbor" OR "harbour" OR "harbors" OR "harbours" OR "haboring" OR "haboring") yields 69,541 results (Set #6)

Finally: TS=("harbor" OR "harbour" OR "harbors" OR "harbours" OR "haboring" OR "haboring" OR "habored" OR "habored") yields the required 80,894 items (Set #7).

As a check we see that #1 NOT #7 yields an empty set. So, lemmatization works.

Using \$ does not turn lemmatization for these terms off. Hence more records are found.

Concretely, TS=(harbor\$ OR harbour\$) yields 80,923 records (Set #8), namely all those in Set #1 plus other ones containing e.g. the term Harbord.

Yet, using ? does turn lemmatization off.

TS=(harbor? OR harbour?) yields 10,692 records (Set #9), while TS=("harbors" OR "harbours") yields 10,663 records (Set #10). Again records containing Harbord are in Set #9 and not in Set #10.

Moreover #9 NOT #10 yields the same result as #8 NOT #1.

Finally using * does turn off lemmatization.

TS=harbor* (set #11) yields 59,605 but not including those with the term harbour (unless also harbor is included, which happens when, to an article written in British English KeyWords Plus added keywords in American English).

Finally, we note that:

TS=(harbor* OR harbour*) yields 81,284 records (Set #12) including all those in Set #1 (#1 NOT #12 is empty) but also including records with the terms Harbourage or Harborth.

Similarly, a search TS=organize (Set #13) yields the same result as:

TS=("organize" OR "organise" OR "organizes" OR "organises" OR "or-

ganized" OR "organised" OR "organizing" OR "organising"),

while TS=organise\$ yields a larger set, including all records of the previous set, but also including records that contain, e.g. the term organizer (and none of the terms of Set #13).

We further note that TS=formula yields the same result as

TS=("formula" OR "formulas" OR "formulae").

The same kind of lemmatization occurs for the prefix TI=. For an author search it is somewhat more complicated: the retrieval result for a search AU=Zhang is the same as that for AU="Zhang", but AU= Zhang J yields the same result as AU=Zhang J*. If one wants to search for Zhang J and not for Zhang JW (for example), then one needs quotation marks: AU="Zhang J".

However, TS=shakespeare yields the same set as TS="shakespeare", while TS=shakespeare\$ yields more than 400 extra records, some containing the plural shakespeare.

CONCLUSION

Lemmatization is probably useful in a large percentage of searches. Yet, occasionally one might want to know the exact content of one's search. Hence I suggest informetricians always to use quotation marks "." which leads exactly to the set one wants to retrieve.

In cases where completeness is essential one may combine (use OR) a lemmatized search with one consisting of search terms between quotation marks. By doing so one does not forget terms that should have been included, e.g. the American or the English spelling variant.

More information about lemmatization can be found e.g. at: http://images.webofknowledge.com/WOKRS53B4/help/WOS/hs_current_limits.html