Foreword

Åström, Fredrik; Danell, Rickard; Larsen, Birger; Schneider, Jesper

Published in:
Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at his 60th Birthday

2009

celebrating scholarly communication studies

a festschrift for olle persson at his 60th birthday

Celebrating Scholarly Communication Studies

A Festschrift for Olle Persson at his 60th Birthday

special volume of the e-newsletter of the
international society for scientometrics and informetrics

vol. 05-S June 2009

Published by **ISSI**

Celebrating Scholarly Communication Studies
A Festschrift for Olle Persson at his 60th Birthday

Special volume of the e-zine of the ISSI, vol. 05-S June 2009

Cover: Balázs Schlemmer
Technical editing: Fredrik Åström & Balázs Schlemmer

Contents

Foreword

The origin of this Festschrift could be traced to a few March days in Hundested, Denmark, in 2001. The event taking place in this small fishing town in the northern parts of Sealand was a PhD course in bibliometrics organized by Professor Olle Persson together with professors Peter Ingwersen and Irene Wormell. Apart from being an excellent course, the Hundested event was also important for bringing together what could be seen as laying the foundation for 'the second generation' of Nordic bibliometrics. As a result, we could see a similar course given in Umeå in 2008, where the core group of teachers were participants of the original course; and also, ending up being the editors of this Festschrift in honour of Olle Persson on his 60th birthday.

Since the 1970s, Olle Persson has been one of the pioneers in Nordic library and information science (LIS) research, analyzing information retrieval, information behaviour and scholarly communication; and of course, the line of research for which he is most closely associated with nowadays: bibliometrics. However, it is not only through his own research that Olle has become a central figure in Nordic LIS, he has also been instrumental in establishing LIS as an academic field of research, where maybe the most tangible effect is the founding of the Nordic Research School in Library and Information Science (NORSLIS) on professor Persson's initiative. And obviously, we must also take into account the number of PhD students Olle has seen through their education as academic supervisor/advisor both in sociology and LIS.

As we all know, Olle Perssons impact is not just limited to Nordic academia and research. Professor Persson has a long list of international publications, many of them highly cited, and it is a list that continues to grow; and in addition to the publishing activities, Olle is also a respected member of the research community, performing duties on editorial boards and program committees as well as e.g. organizing the 10th ISSI conference in Stockholm in 2005.

One particular aspect of Olle Perssons work that cannot go unmentioned is the Bibexcel toolbox for bibliometricians (see also the contributions by Meyer & Glänzel, and Persson, Danell & Schneider in this Festschrift): a software of huge impact for many of us doing bibliometric analyses; but also something that reflects both Olle's flexible mind and generous nature through how the software has been developed over the years – many times by Olle adding functionalities at the request of other users. Apart from Olle Perssons strengths as a scholar and a teacher, we should also mention some of those personal traits making Olle not only a good academic, but also a well esteemed colleague and a good friend. On one hand, we have that relentless energy and curiosity that never cease to amaze us, pushing Olle's activities forward as well as those by us being fortunate enough to work with him. On the other hand, we also have the strong drive from Olle to do things for other people.

The festschrift

A few words on the contributions in celebration of Olle Persson. A rare feature in a Festschrift is an article by the one being celebrated. However, when this volume started to take shape, one idea we felt was important was to include a formal description of the Bibexcel software, something that has been missing for a long time; and, something making it easier for us to acknowledge our intellectual debt to Olle by having a document to cite when having used Bibexcel. But, instead of having Olle write it himself, as a birthday gift to him, we decided to write it, but obviously, have him being first author.

The international and cooperative nature of research being both analyzed and appreciated by Olle, is reflected in this Festschrift bringing together contributions from Spain, the UK, Belgium, Norway and the US, as well as Sweden, Denmark and – to some extent – Australia. The Bibexcel paper is the obvious starting point, followed an investigation of research evaluation in the fields of the humanities and social sciences by Isabel Iribarren-Maestro, María Luisa Lascurain-Sánchez and Elias Sanz-Casado. In the second paper, Martin Meyer and Wolfgang Glänzel analyze the impact of professor Persson's software Bibexcel on the mapping of research fields.

Ronald Rousseau addresses a new topic in his pilot study on the impact of highly influential editorials in scientific journals, identifying problems related to the definition of what an editorial is, as well as how they are indexed in the Web of Science databases. The issue of research evaluation is revisited by Gunnar Sivertsen, also addressing problems related to how the social sciences and humanities can be assessed using bibliometric methods; and how that relates to differences in publication patterns in different research fields.

As well as with Meyer and Glänzel's article, Mike Thelwall address the impact of Olle Persson, not by looking at his scholarly articles but by looking at his online activities, such as the Inforsk Research Group website and the Bibexcel software. In the last article, Howard White combines relevance theory from linguistic pragmatics and ideas from informetrics and information retrieval for producing and interpreting pennant diagrams; and as an example, uses co-citation analyses on Olle Persson and August Strindberg.

Dear Olle!

With this Festschrift, we want to celebrate your 60th birthday, to show our appreciation for you as colleague and friend, as well as mentor and teacher. We are many that have you to thank for a lot of what we are doing nowadays, something that is reflected not only in us wanting to present you with this volume, but also through the impact evident in the articles analyzing your work here.

Happy 60th Birthday!
June 15, 2009
Aalborg, Copenhagen, Umeå and Sydney

FREDRIK ÅSTRÖM
Lund University Libraries,
Head Office, P.O. Box 134, SE-22100 Lund (Sweden)
University of Technology Sydney, Faculty of Arts and Social Sciences,
P.O. Box 123, Broadway, NSW 2007 (Australia)
E-mail: fredrik.astrom@lub.lu.se

RICKARD DANELL
Umeå University, Dept. of Sociology, SE-90187 Umeå (Sweden)
E-mail: rickard.danell@soc.umu.se

BIRGER LARSEN
Royal School of Library and Information Science,
Birketinget 6, DK-2300 Copenhagen S (Denmark)
E-mail: blar@db.dk

JESPER WIBORG SCHNEIDER
Royal School of Library and Information Science,
Fredrik Bajers Vej 7K, DK- 9220 Aalborg Ø (Denmark)
E-mail: jws@db.dk

How to use Bibexcel for various types of bibliometric analysis

Olle Persson[1], Rickard Danell[1] and Jesper Wiborg Schneider[2]

[1] Department of Sociology, Umeå University, SE-90187 Umeå, Sweden
[2] Royal School of Library and Information Science, DK-9220, Aalborg

Introduction

Bibexcel is a versatile bibliometric toolbox developed by Olle Persson. In Bibexcel it is possible to do most types of bibliometric analysis, and Bibexcel allows easy interaction with other software, e.g. Pajek, Excel, SPSS, etc. The program offers the user high degree of flexibility in both data management and analysis and this flexibility is one of the program's real strengths. It is, for example, possible to use other data sources than Web of Science, and Bibexcel can in fact deal with data other than bibliographic records. If the user only learns the basic file structures that Bibexcel requires it possible to import many different types of data. However, flexibility has its price and the flexibility may initially cause new users to perceive it as difficult to use. We therefore find this festschrift an appropriate forum to describe Bibexcel's basic functions, and we will describe these functions by analyzing data consisting of Web of Science articles that cite Olle Persons scientific publications.

The chapter is structured in four main sections. In the first section we describe how data downloaded from Web of Science must be restructured. In the second section we will take a closer look at the OUT-file. Bibexcel produces several types of files; in fact every procedure will give the user a new file. However, the OUT-file is always created first, and it is this file that is the starting point for the analysis you want to do in Bibexcel. In the third section we will give a brief description of basic analytical functions available in Bibexcel. In the last section we will describe how to export files to Pajek in order to do visualizations. The aim of this chapter is to introduce readers to how to use Bibexcel. We will assume that the readers have some basic knowledge of bibliographic data and basic bibliometric techniques. It should also be noted that Bibexcel includes far more features than described in this chapter. However, it is our hope that our basic description will make it possible for the interested reader to acquire enough knowledge to start using Bibexcel, and after some experimentation be sufficient self reliant to figure out how to use functions not described in this chapter.

How to prepare and import data

As stated above, it is possible import many different data formats in Bibexcel, but in this section we will only to comment on how to import data from Web of Science. The data imported from Web of Science should be saved as plain text. The plain text file downloaded from Web of Science needs to be restructured before it can be imported into Bibexcel. The restructuration of data consists of two simple steps. First, it is necessary to insert carriage return in the text file we want to import in Bibexcel. Carriage return can be inserted in the file in two different ways; either we open the file in word and re-save it as a text file, or we select from Bibexcels menu:
Edit doc file->Replace line feed with carriage return

The next necessary restructuring of the file downloaded from the Web of Science is to convert the bibliographic records to a DIALOG format. We do this by first selecting the file with the extension *.tx2 and then choose the following option from the menu:
Misc->Convert to dialog format->Convert from Web of Science

The procedure tells Bibexcel to create a file with the extension *. doc (we henceforth refer to this as DOC file). After completing these steps the text file we downloaded from the Web of Science is ready to be used in Bibexcel and we can start analyze the data. However, before we start, we may want to familiarize our self with the structure of the DOC-file. At least it is necessary to be familiar with the structure of the bibliographic records in the DOC-file. Bibexcel keeps track of where the bibliographic records begins and ends by looking for a double-spike, that is | |. Each record is composed of several bibliographic fields and Bibexcel keeps track of where the bibliographic fields begins by field tags. For example, the field tag for the author field is "AU". It is important to keep track of field tags, because we usually have to tell Bibexcel which bibliographic fields we want to work with. Each bibliographic field ends with a single spike, i.e. |. In bibliographic fields with multiple units, the units are separated from each other with some delimiter. For most bibliographic fields the field delimiter is a semicolon. However, there are other delimiters and it is necessary to tell Bibexcel how the bibliographic field is delimited.

Creating an OUT-file and calculating frequency distributions

We made the file containing the bibliographic records readable in Bibexcel, and we are now ready to make an OUT file. Making an OUT-file is always the first step when analyzing bibliographic data with Bibexcel. When we make the OUT-file we start by decide which bibliographic field the OUT-file will be constructed from. We tell Bibexcel which bibliographic field the OUT-file will be constructed from by entering the field tag in the box marked "Old tag", e.g.

if we want the OUT file to be based on the author field we write AU in "Old tag".

Next we select the DOC-file from the file managing system by clicking on it. We must know how the field is delimited, and we select an option for how the units in the field are delimited from the scrollbar marked “Select field to be analyzed”. In our example we will make an OUT-file based on the author field and in this field each co-authors are separated from each other with a semicolon. We therefore select the option “Any ; separated field”.

After we selected the DOC-file, and selected field delimiter, and typed AU in the box marked “Old tag” we press the button labeled "Prep." Bibexcel now creates an OUT-file based on the author field, and each line in the OUT-file will be matched by a unique authorship as a unique author holds. The OUT-file's structure is simple and it is important that we familiarize ourselves with it. The OUT-file we produced has the following structure (Table 1):

Table 1 The structure of the OUT-file

Document identification number	Authors
1	Levitt JM
1	Thelwall M
2	Hsu PY
2	Shiau WL
2	Su YM
2	Yang SC
3	Anegon FD
3	Guerrero-Bote VP
3	Olmeda-Gomez C
3	Ovalle-Perandones MA
3	Perianes-Rodriguez A
….	….
….	….
388	Stefaniak B
389	Vlachy J
390	Ellis D

The left column of the OUT-file consists of a document identity number, and the right column the authors. In Table 1 we see that in total there are 390 documents citing Olle Persson’s publications and the first document in the DOC-file has been co-authored by JM Levitt and M. Thelwall. Regardless of what bibliographic field we select the OUT-file will have this structure. It can be of interest to note that the OUT-file is a tab-delimited text file, and like all Bibexcel files, it can be imported in Excel, or other statistical software.

Frequency distributions

Depending on what bibliographic fields we have chosen as a unit when we created the OUT-file, the frequency calculation function in Bibexcel offers many different options. For example, if the OUT-file consists of cited document Bibexcel can make a substring search and only count a specified part of the cited document, e.g. cited journal or cited author. In Bibexcel, we can also choose between two counting methods when we ask Bibexcel to count units in the OUT-file: "whole counts" and "fractional counts". If we chose to check the box marked "Fractionalize" Bibexcel will change counting methods to "fractional counts" and if this boxed is left unchecked the counting method will be "whole counts". The method of fractional counting is easy to understand. For example, if a document is co-authored by two authors each author attributed half an article and if the document has three co-authors each author will be attributed a third of the article, etc.

In our example the OUT-file consists of authors who have cited at least one of Olle Persson's publications, and we want to know how many times the authors have cited Olle Persson. To get Bibexcel to calculate the distribution of authorships we select the OUT-file from the Bibexcel's file managing system (under the caption "Select file here"). Next we must tell Bibexcel whether it should count "whole strings" (i.e. the whole row in the OUT-file) or some predefined part of the text string. In our example the rows OUT-file consists of author names and Bibexcel cannot do a substring search in an author name. Since we want Bibexcel count the whole author name we select the option "Whole String" from the scrollbar under the caption "Select type of Unit". If we want the list of authors sorted descending by frequency, we click in the box marked "Sort descending" and if we want to change counting method to "fractional counts" we click in the box marked "Fractionlize". We start the counting procedure by pressing the button marked "Start", and Bibexcel creates a frequency distribution and saves it in a file with extension *.cit.

Table 2 Authors citing at least one of Olle Persson's publications

Author	Whole counts	Fractional counts
Persson O	15	7,999
Leydesdorff L	14	9
Glanzel W	13	7,416
Meyer M	13	9,833
Melin G	12	7,916
Zitt M	8	3,332
White HD	7	5,5
Rousseau R	6	2,366
Gomez I	6	2,199
Zuccala A	6	4,5
Cronin B	6	4,333
Moya-Anegon F	5	1,999
Morris SA	5	2,083
Bassecoulard E	5	2,166
Herrero-Solana V	5	1,749

Creating a new OUT-file

In some situations, it is necessary redefine the units in the OUT-file and to make a new OUT-file. If we, for example, are interested in examining which journals that has been cited, Bibexcel enable us to make a new OUT-file in which an OUT-file containing cited documents are reduced to a OUT-file containing cited journals, i.e. Bibexcel removes all references that are not published in a scientific journal, and keeps only the name of the cited journals in the new OUT-file. I should be noted, in this procedure Bibexcel assumes that a cited document with a volume number is published in a journal. This definition is not perfect and we usually need to do some editing of the new OUT-file. Alternative options to make a new OUT-file is listed in the scrollbar under the caption "Select type of unit". That is, in the same scrollbar that we use to select the type of unit when we created the CIT-file.

To create a new OUT-file we start by selecting the "old" OUT-file in the box for file management. We continue by selecting a type of unit in the scrollbar under the caption "Select type of unit". In our example, we will select the option "Cited journal". Next we decide if we want to Bibexcel to eliminate duplicates, i.e. remove identical units with the same document identification number. Since the units in our new OUT-file will be cited journals we should consider this option; it is possible that several documents in the reference list has been published in the same journal. Duplicate units in the OUT-file will cause problems for some types of co-occurrence analysis, e.g. if we use a MDS algorithm we do not want loops in the matrix. If we click in the box marked "Remove duplicates" all units with the same document identification number will be unique. Next we tell Bibexcel to make a new OUT-file by clicking in the box marked "Make new out-file" and pressing the button labeled "Start". The new OUT-file will have the extension *. oux. The OUX-file has the same structure as the OUT-file, and we can use the OUX-file in the same way we use the OUT-file, e.g. we can tell Bibexcel to calculate how many times the journals has been cited. To illustrate the effect of removing duplicates from the OUX-file Table 3 displays the citation distributions over journals with duplicates removed and with duplicates included.

Table 3 Scientific journals usually used when citing Olle Persson publications

Journal	No. Citations (no duplicates)	No. Citations (all citations)
SCIENTOMETRICS	344	2691
J AM SOC INFORM SCI	188	988
RES POLICY	151	472
SCI TECHNOL	110	127
SCIENCE	103	198
SOC STUD SCI	101	217
J INFORM SCI	89	179
INFORM PROCESS MANAG	86	201

Journal	No. Citations (no duplicates)	No. Citations (all citations)
J AM SOC INF SCI TEC	85	328
J DOC	79	246
NATURE	56	110
SCI PUBL POLICY	54	78
ANNU REV INFORM SCI	50	111
RES EVALUAT	49	73
AM SOCIOL REV	46	87
AM PSYCHOL	42	60
AM J SOCIOL	37	64
P NATL ACAD SCI USA	37	63
LIBR TRENDS	37	65

Producing a data matrix for export to a statistical software

A useful feature in Bibexcel is the one that enables us to produce data matrices for export to statistical software. In Bibexcel a data matrix is created by adding variables to an OUT-file, and the variables are created by selecting fields in the bibliographic record. However, there is a pitfall that Bibexcel users should be aware of. We must remember that the first variable in the matrix, i.e. the unit in the OUT-file, must be present in all bibliographic records. If the bibliographic field, chosen as unit when the OUT-file was made, is missing in some records Bibexcel will not be able to create a data matrix. It thus makes sense to choose a field that is always available, e.g. the field containing the journal name (SO). If the units in the OUT-file are present in all bibliographic records, it is no problem to add fields to the OUT-file that are missing in some bibliographic records; Bibexcel will write "No field" as compensation for the lost variable value.

If we have created an OUT-file, that contains all the document identification numbers, the procedure for creating a data matrix quite simple. We begin by selecting the OUT-file (or an OUX-file, or some other file with the same structure as an OUT-file), and enter a field tag in the box marked "Old tag". If we want to add “times cited” to the OUT-file as a variable we write TC in box labeled "Old tag". By entering the field tag in the "Old tag" we tell which field we want Bibexcel to use when it adds a new variable to our data matrix. Next, we press the button labeled "Add fields to units”. The file created by Bibexcel has the extension *. jn1. If we want to continue to add variables to our matrix we select the JN1-file and enter a new field tag in "Old tag" and press "Add fields units”. This makes Bibexcel creates a file with the extension *. jn2. We continue by selecting the JN2-file and repeat the procedure and Bibexcel creates a JN3-file. We continue until the matrix contains sufficient information. Since JN *- file is a tab-delimited text file, it can be imported into Excel or any other statistical software for further analysis.

Table 4 Example of a data matrix produced by Bibexcel

Document identification number	Times cited	Publication year
1	0	2009
2	0	2009
3	0	2009
4	0	2009
5	0	2008
6	0	2009
7	0	2008
8	0	2009
9	0	2009
…	…	…
389	3	1987
390	3	1986

Co-occurrence analysis and the COC-file

The present section illustrates how to perform some of the most essential bibliometric analyses by use of Bibexcel. The previous Bibexcel menus, *File*, *Edit doc* file, and *Edit out-file*, have focused upon cleaning and setting up data for analysis. The menu *Analyze* presented here contains a number of specialized functions that allow us to perform analyses of citation networks and, perhaps most important, a range of different co-occurrence analyses. We cannot demonstrate all the functions in the *Analyze* menu. We will therefore focus upon co-occurrence analysis and *how* to prepare your data and how to perform co-occurrence analyses. We demonstrate the intrinsic functions and point out the small differences in performing co-citation analysis, bibliographic coupling, co-word analysis, and co-author analysis. Finally, we look at some alternatives in the *Analyze* menu that enable some more advanced multivariate analyses.

Co-occurrences

A bibliographic record consists of a number of fields used to index the actual text, its subjects and descriptive data. As demonstrated above, when working with Bibexcel we usually transforms our initial data to the *Dialog*-format, more specific the format for *Science Citation Index*®. Common data between records are thus structured in univocal metadata fields, such as publication titles in the title field, authors in the author field, and references in the reference field. A co-occurrence relation in a bibliographic record usually means the mutual occurrence of two units in the same metadata field. Hence, when words *x* and *y* appear together in the title field, or when authors *z* and *w* appear together in the author field. Obviously, *one* co-occurrence relation between two units is trivial. What is interesting, on the other, is whether a co-occurrence relation between two units is frequent over a number of records, for example that the same title words *x* and *y* appears together in a number of records, or the same pair of authors *z* and *w* also appear together in a number records – this is in principle a co-occurrence analysis. Co-occurrence analysis is therefore the study of mutual appearances of pairs of units over a consecutive number of bibliographic records. With this in mind, we will now illustrate how we prepare and perform co-occurrence analysis in Bibexcel.

Preparing data and files for co-occurrence analysis

Basically, the co-occurrence routine in Bibexcel match pairs of units, extracted from the same metadata field, within and across the records in the DOC-file. In order for the routine to match pairs of units, we need to indicate 1) on what file the matching routine is to be performed; and 2) what units should be matched.

The 'workhorse' in Bibexcel is the OUT-file. Most analytical routines are performed on the OUT-file. Remember that the OUT-file is basically an extraction of units from one metadata field across the records in the DOC-file. Therefore, the unit of analysis in the OUT-file defines the type of co-occurrence analysis. For example, an OUT-file that lists the individual authors from each record in the DOC-file, would be the basis for a co-author analysis. The matching routine used to match pairs of units must therefore be performed on the OUT-file. It is the unit pairs in the individual documents and their frequency cross all documents that must be generated.

Notice, many individual units will have very low frequencies. Such units are often unimportant in co-occurrence analyse as their mutual relationships will be trivial due to low frequencies. It is therefore a very good idea to use individual frequency as an inclusion criterion for the analysis. Further, such a criterion also speeds up the generation of co-occurrence pairs, since this can be a resource demanding routine depending on the number of units to match. Obviously, other criteria can also be used to reduce the number of units for the analysis. In order to select the units to be included and matched, we need the CIT-file. Remember, the CIT-file is a frequency file of the units represented in the OUT-file. What we need to do is to show the CIT-file in *The List*, which is done by marking the CIT-file in *Select file here* and then press the *View file* button.

Let us illustrate the preparation procedure. The first thing to do is to actually select the units to be included in the analysis. Mark the CIT-file in *Select file here* and show the CIT-file in *The List* by pressing the *View file* button. Next, we must mark the units from the CIT-file (shown in *The List*) that we want to study in the co-occurrence analysis. Pointing the mouse to one unit will mark it. To mark several consecutive units, mark the first, then press *SHIFT* and mark the last unit; to mark a number of non-consecutive units, press *CTRL* and the use the mouse to select units. If you regret the selection of a unit, press *CTRL* and use the mouse again on the chosen unit. Notice if you need to view and select many units from *The List* press the *View whole file* and scroll down and select units. Below is an example of a CIT-file that contains frequencies of cited references (represented by author and publication year):

Table 5 The CIT-file

Frequency	Cited reference
99	LUUKKONEN T, 1992
79	KATZ JS, 1997
74	MELIN G, 1996
59	LUUKKONEN T, 1993
56	WHITE HD, 1981
50	PERSSON O, 1994
45	NEWMAN MEJ, 2001
44	BEAVER DD, 1979
42	BEAVER DD, 1978
42	WHITE HD, 1998
99	LUUKKONEN T, 1992
79	KATZ JS, 1997

To select all 10 references mark the first unit, press *SHIFT* and mark the last unit. Notice, you mark the whole line, including the frequency number.

We are now ready to commence the actual co-occurrence analysis. This is done by the following procedure from the Bibexel menu:
Analyze -> Co-occurrence -> Select units via listbox

This routine removes non-selected units from the *The List* so that only the ones selected for the co-occurrence analysis are kept.

The next step is to identify co-occurrence relations between the selected units, i.e., the actual co-occurrence analysis. This step is done in two tempi. First, we need to indicate what file the matching routine is to be performed on, as indicated above. This is always the OUT-file. We therefore mark the OUT-file in *Select file here*. Notice, we only mark the OUT-file, we do not show it in *The List*, where we already have the units selected by the previous operation. Consequently, we have marked the OUT-file in the *Select file here* and the selected units are in *The List*; only the first 17 are visible. The next move is to run the matching routine on the OUT-file, which is done by the following procedure from the Bibexcel menu:
Analyze -> Co-occurrence -> Make pairs via listbox

A question pops up immediately after activating the routine, asking whether one wishes to include individual frequencies for the units in addition to co-occurrence frequency in the output. Most often, if the purpose of the analysis is a mapping of some sort, such frequencies should be left out. The outcome of the co-occurrence routine is the COC-file (abbreviation for co-occurrence), examples without and with individual frequencies is shown below.

Table 6 The COC-files

COC freq.	Cited reference #1	Cited reference #2	COC freq.	Cited reference #1	Cited reference #2
42	KATZ JS 1997	MELIN G 1996	42	KATZ JS 1997$79	MELIN G 1996$74
42	BEAVER DD 1978	BEAVER DD 1979	42	BEAVER DD 1978$42	BEAVER DD 1979$44
34	WHITE HD 1981	WHITE HD 1998	34	WHITE HD 1981$56	WHITE HD 1998$42
33	PERSSON O 1994	WHITE HD 1981	33	PERSSON O 1994$50	WHITE HD 1981$56
33	BEAVER DD 1978	LUUKKONEN T 1992	33	BEAVER DD 1978$42	LUUKKONEN T 1992$99
31	BEAVER DD 1979	LUUKKONEN T 1992	31	BEAVER DD 1979$44	LUUKKONEN T 1992$99
30	KATZ JS 1997	LUUKKONEN T 1992	30	KATZ JS 1997$79	LUUKKONEN T 1992$99
28	MELIN G 1996	NEWMAN MEJ 2001	28	MELIN G 1996$74	NEWMAN MEJ 2001$45
Example of COC-file, where individual frequencies are not included			*Example of COC-file, where individual frequencies are included*		

The COC-file contains actual co-occurrence frequencies between the units chosen for the analysis. The co-occurrences frequencies are based on matching *all* possible parings of units in the OUT-file, to establish how many times they appear together across the records.

Opportunities with COC-file

There are several opportunities from here. The COC-file can very easily be transformed into an input file for network visualization in Pajek. This is demonstrated in the next section. The COC-file is also the input for the special clustering algorithm in Bibexcel, called *Persson's Party Clustering*. You simply mark the COC-file in *Select file here* and then proceed with:
Analyze -> Co-occurrence -> Cluster pairs

The outcome is a PER-file that contains information about the number of clusters, their members, and links within and between clusters. Finally, the COC-file can also be used as the basis for construction of square symmetric matrices. This is done in two steps. First, we need to indicate which units that will go into the matrix. Most often this is done by marking the COC-file in *Select file here* and then proceeds with:
Analyze -> List units in pairs

The outcome is a CCC-file. To proceed with the matrix generation, mark the CCC-file in the *Select file here*, and from the *Analyze* menu choose:
Analyze -> Make a matrix for MDS etc.

You have the choice of generation a square symmetric matrix or its constituent lower left part. The latter is the input for the MDS SYSTAT algorithm that is compatible with Bibexcel, but not an integrate part of the software. If you have access to SYSTAT, continue with the routines *Make a map/SYSTAT cmd file* and *Show map* to produce an MDS mapping of the matrix. Otherwise export the matrix to Excel, SPSS, UCINET or other softwares containing MDS routines.

Different co-occurrence analyses

Above we have described the intrinsic procedures of co-occurrence analysis in Bibexcel when using the *Analyze -> Co-occurrence* routine. Table x below outlines different types of co-occurrence analyses, what units should be in the OUT-file to produce them, how to run them, their outcome files, and some of their special characteristics.

Table 7 Matrix of co-occurrence analyses

Type of co-occurrence analysis	OUT-file must be based on this metadata field	Bibexcel menu:	Special characteristics
Co-citation analysis	CR or CD; i.e., the cited reference string, or part of if, such as cited author or cited work.	*Analyze -> Co-occurrence -> Make pairs via listbox* *Outcome = COC-file*	The focus is upon mapping relations between cited references
Bibliographic coupling	CR or CD; i.e., the cited reference string, or part of if, such as cited author or cited work.	*Analyze -> Shared units* *Outcome = COU-file (similar format as the COC-file)*	The focus is upon mapping relations between citing publications (the records in the DOC-file)
Co-author analysis	AU or AF, and	*Analyze -> Co-occurrence -> Make pairs via listbox* *Outcome = COC-file*	
Co-word analysis	TI (title words) , AB (words from abstract), DE (descriptors); SC Subject categories) etc.	*Analyze -> Co-occurrence -> Make pairs via listbox* *Outcome = COC-file*	In principle every field, if it makes sense, can be used; even 'homemade' fields can be made if for example data are imported from other sources, see Schneider (2004).

For example, to do a document co-citation analysis, you need to generate an OUT-file with cited references and then proceed as described above. The same goes for a co-author analysis, where you generate an OUT-file containing authors extracted for example from the AU field in the DOC-file. Co-word analysis actually covers a number of co-occurrence analyses, depending on the units in the OUT-file, for example title words, descriptors, subject categories etc. – you decide.

The only deviance from the intrinsic procedure is bibliographic coupling. Bibliographic coupling is called *Shared units* in Bibexcel and the routine is performed in the following way. Mark the OUT-file in *Select file here* and
Analyze -> Shared units

The outcome is a COU-file that contains pairs of documents and their common share of units (references). Notice that documents are indicated by the record number. You may want to add labels to these numbers, for detailed description see the Help file in Bibexcel. The COU-file is in principle the same as a COC-file and can be used as input for clustering and matrix generation, as described above. [

More advanced co-co-occurrence analysis: The MA5-file

Sometimes we need to do more elaborate multivariate analyses of co-occurrence relations. For such purposes, Bibexcel provides an extremely valuable tool, *Docs and units matrices*. This routine creates a classical data matrix, where rows correspond to documents (records) and columns the chosen units for the study.

Creating Pajek files in Bibexcel

In this section, we shall describe how to create Pajek files in Bibexcel. However, we will not give a detailed account for the many functions in Pajek. Pajek is freely available on the Web, and for those interested in this excellent software should visit Pajek Wiki (http://pajek.imfm.si/doku.php). This resource contains both the program and a manual.

In Bibexcel, we can create three types of Pajek files: network files (NET-file), vector files (VEC-file), and files containing information about the cluster membership (CLU-file). In this section, we shall describe how to produce these files in Bibexcel and we will use the files to visualize Olle Persson's position in a document space, or if you like Olle Persson's intellectual contexts.

If we want to create files for Pajek in Bibexcel we must follow a specific sequence. We always start by creating the network file (NET-file), and a NET-file is created from a COC-file. We start by selecting the COC-file, and from the Bibexcel menu we choose:
To Pajek -> Create net-file from coc-file

Since a network may be a directed or undirected graph, Bibexcel will ask us "Do your data have directed arcs, like A cites B". Since the COC-file used in this example is the result of a co-citation analysis we do not have directed arcs and we answer no to the question. Bibexcel completes the process and creates a file with the extension *.net, which contains the co-citation network. After we made the NET-file, we can continue to create VEC-files and CLU-files in any order. The reason we need to do NET-file first is that the process that creates NET-file also creates a file with extension *. Vel, which Bibexcel need to make VEC and CLU files.

To make a VEC file, we use a CIT file. In our visualization of Olle Persson's intellectual contexts we also want to show which documents that are most cited. We will therefore create a vector file that contains information about the co-cited papers citation frequency. To do this VEC-file we start by selecting the CIT-file that contains citation frequencies, and then we select:

To Pajek -> Create vec-file

It is possible to use other files than CIT files when we want to make a VEC file. What is important is that the file has the same structure as CIT-file, i.e. a tab-delimited text file where the first column consists of some values for the vector and the second column the name of the nodes included in the NET file.

A third type of Pajekfilen we can create in Bibexcel is the CLU-file, which contains information about the partitions. You can create partitions in Bibexcel with an algorithm developed by Olle Persson (Persson's party cluster). However, the CLU-file can be based on any type of partition principle, as long as the file we tell Bibexcel to use has the right structure. In our example, we shall use Olle Persson own algorithm and for more details of the algorithm see Persson (1994).

To partition the co-citations matrix with Olle Persson algorithm we select COC-file and from the menu we choose:

Analyze -> Co-occurances -> Cluster pairs

Bibexcel will create three files, each containing information about the clusters created from the COC-file. The files has the extension *. pe2, *. pe3, and *. per. The file we need to make a CLU-file is the PE2-file. We select the PE2-file and from the menu we choose:

To Pajek -> Create clu-file

Importing NET, VEC and CLU files in Pajek is simple. The NET file we open as "Networks", the VEC-file we open as "Vectors" and the CLU file we open as "Partitions". After we have opened the files in Pajek, we choose the following option from the Pajek menu:

Draw -> Draw-Partion-Vector

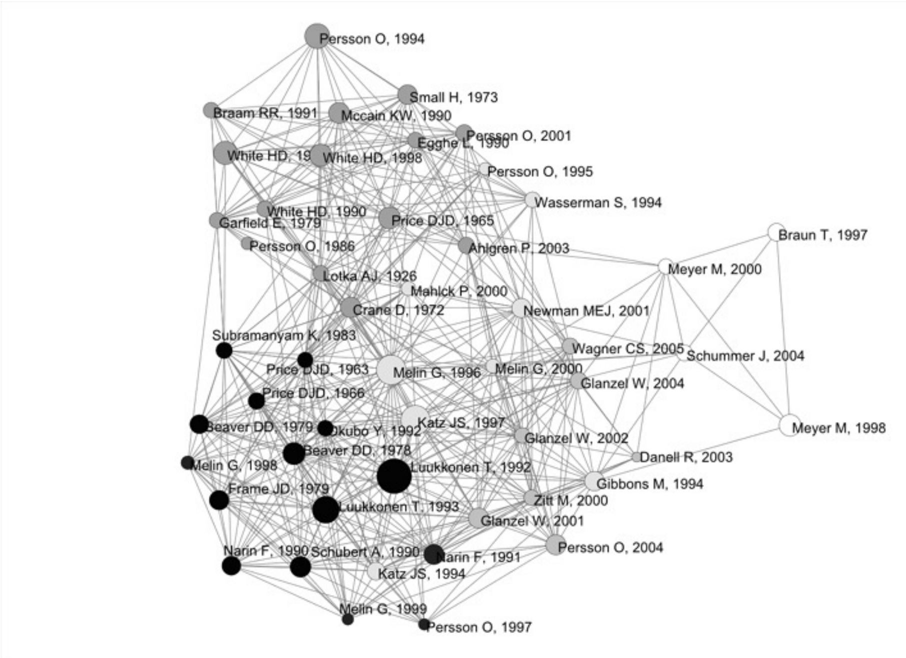

Figure 1 Olle Persson intellectual context

The map displayed in Figure 1 has been created with Pajek. The co-citation map shows the context in which Olle Persson scholarly works have been used. The documents are represented by the first author and publication year. It should be noted that many of Olle Person's publications are represented by other authors, and the most cited of these are Luukkonen T, 1992, Luukkonen T, 1993 and Melin G, 1996, which is Luukkonen, Persson & Sivertsen (1992), Luukkonen, Persson & Tijssen (1993), and Melin & Persson (1996). The cluster algorithm produced six clusters, and we can aggregate them into three main intellectual themes. In the upper part of the map, we find publications primarily used in an information science context. Most cited of these is Persson (1994), which is Olle's analysis of the intellectual base of *Journal of the American Society for Information Science*. It is in this article, which is an author co-citation analysis, that Olle presents his algorithm for "party clustering". The article is highly co-cited with Henry Small's classic article and several articles by White and McCain. The main theme of this cluster is obviously co-citations analysis. In the same cluster, but further down the map, we find Olle Persson article on all author co-citation analysis, a proposed solution to the first author problem in traditional author co-citation analysis. Another article by Olle Persson in this cluster is Olle's article on Online bibliometrics (Persson 1986). In the middle and the lower left part of the map we find publications that address issues of research collaboration and science internationalization. A close examination of the co-cited papers in the smaller cluster in the map's center reveals that the orientation of these studies is on formation of social networks, while the larger cluster deals with issues relating to the increasing share of co-authored articles and internationalization of science. On the map's right side, we find articles dealing with research evaluation and technology transfer. The map gives us an insight into the breadth and importance of Olle Persson scientific achievement.

References

Danell, R. and O. Persson (2003). "Regional R&D activities and interactions in the Swedish Triple Helix." Scientometrics 58(2): 205-218.

Glanzel, W., R. Danell and O. Persson (2003). “The decline of Swedish neuroscience: Decomposing a bibliometric national science indicator.” Scientometrics 57(2): 197-213

Luukkonen, T., O. Persson and G. Sivertsen (1992). "Understanding Patterns of International Scientific Collaboration." Science Technology & Human Values 17(1): 101-126.

Luukkonen, T., R. J. W. Tijssen and O. Persson (1993). "The Measurement of International Scientific Collaboration." Scientometrics 28(1): 15-36.

Mahlck, P. and O. Persson (2000). "Socio-bibliometric mapping of intra-departmental networks." Scientometrics 49(1): 81-91.

Melin, G. and O. Persson (1996). "Studying research collaboration using co-authorships." Scientometrics 36(3): 363-377

Melin G. and O. Persson (1998) “Hotel cosmopolitan: A bibliometric study of collaboration at some European universities”, Journal of the American Society for American Society for Information Science, 49(1): 43-48

Melin, G., R. Danell and O. Persson (2000). "A bibliometric mapping of the scientific landscape on Taiwan." Issues & Studies 36(5): 61-82.

Meyer, M. and O. Persson (1998). "Nanotechnology - Interdisciplinarity, patterns of collaboration and differences in application." Scientometrics 42(2): 195-205.

Persson, O. (1986). "Online Bibliometrics – A Research Tool for every Man." Scientometrics 10(1-2): 69-75.

Persson, O. (1994). "The Intellectual Base and Research Front of JASIS 1986-1990." Journal of the American Society for Information Science 45(1): 31-38.

Persson, O. and M. Beckmann (1995). "Locating the Network of Interacting Authors in Scientific Specialties." Scientometrics 33(3): 351-366.

Persson, O. (2001). "All author citations versus first author citations." Scientometrics 50(2): 339-344.

Persson, O., W. Glanzel and R. Danell (2004). "Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies." Scientometrics 60(3): 421-432

Schneider, J.W. (2004). Verification of bibliometric methods' applicability for thesaurus construction.

PhD dissertation. Aalborg. Department of Information Studies, Royal School of Library and Information Science, 2004. xiii, 356 p. plus Appendix volume, 125 p.

Addresses of congratulating authors:

RICKARD DANELL
Department of Sociology, Umeå University,
SE-90187 Umeå, Sweden
E-mail: rickard.danell@soc.umu.se

JESPER WIBORG SCHNEIDER
Royal School of Library and Information Science,
DK-9220, Aalborg
E-mail: jws@db.dk

The Use of Bibliometric Techniques in Evaluating Social Sciences and Humanities

Isabel Iribarren-Maestro, María Luisa Lascurain-Sánchez & Elias Sanz-Casado

Carlos III University of Madrid, C/Madrid, 126. Getafe 28903 Madrid (Spain)

Research in the social sciences and humanities: Evaluation Problems

Research is one of the most important tasks being performed in the world today. It is difficult, if not impossible, to understand the future of mankind if divorced from scientific activity. The challenges that human beings face can only be dealt with through knowledge, technology and innovation. To date, human activity on the planet has been closely linked to the development and use of knowledge obtained through the hard sciences. This means that research, and especially technology, have been defined along fairly narrow lines. However, new trends more in line with the times, in which sustainability plays an important role, are causing scientific activities to be undertaken from a more cross-disciplinary perspective, with the participation of knowledge acquired from different areas. In this new paradigm, the social sciences and humanities are contributing a type of knowledge that is more immediate to ever-changing, more flexible needs and realities.

The active participation of all fields in knowledge-generating processes must be evaluated, and social science and the humanities are no exception. Such evaluation may be performed using bibliometric indicators, as is done in other sciences, given that the only requirement is that researchers must produce scientific knowledge that can be quantitatively analyzed. However, certain characteristics of researchers in these fields are obstacles to measurement and assessment using the same criteria as for scientists working in the pure, experimental and technical sciences, also know as “hard sciences”. Indeed, since their publication habits differ considerably from those of other groups , most of their research cannot be retrieved from the databases normally used to obtain information that can be bibliometrically analyzed.

The habits and characteristics of scientists working in the social sciences and humanities have been analyzed and described in both bibliometric and user studies. These studies began to shed light on certain characteristics in both publications and in the use of information to generate new knowledge. As the characteristics in question had been unknown until then, the features characterizing other scientific groups were attributed to these scientists by default. The results of bibliometric and user studies portrayed a previously unimagined reality in which the profiles of the research habits of scientists working in the social sciences and humanities began to be defined, and the information centres most suitable for meeting their information needs began to be designed (Brittain, 1979; Siatri, 1999).

In the case of humanists, the origin of these studies was the project launched in 1976 by the Centre for Research in User Studies (CRUS) and funded by the British Library, whose objective was to *explore user information needs and behaviour,* identifying such aspects as their limited ability to work in teams or the types of documents they use. Research on the habits of social scientists dates back farther than the studies on the humanities, to the late 1960s. These studies were triggered by both the libraries' lack of knowledge about this group and by the interest of professional social science associations in adapting the content of their courses and programmes to the information needs of their member researchers. Furthermore, the surge in development that took place in some social sciences, such as economics and psychology, made it necessary to create new information centres to meet the needs of scientists in these fields. To this end, studies were conducted to ensure that these centres would be as well suited to user needs as possible.

Now, more than thirty years after these beginnings, an intense need has arisen to understand and evaluate the scientific production of social science and humanities researchers, for which it is essential to define the features that characterize them and set them apart from researchers in other areas of knowledge. Indicators with which to efficiently and precisely determine what resources have been invested in carrying out their scientific activity must also be developed.

The publication of research results is fundamental for scientists, as it enables them to disseminate their activities among the scientific community, which can then compare and validate the findings. Despite the imperfections widely acknowledged, this represents an important hurdle to be overcome in research validation and concomitant inclusion in the scientific *acquis*. However, in the social sciences and humanities, publications are often poorly reflected in national and international databases, partly because of the characteristics of the sources in which they are published, and partly because of the limited resources that the producers and distributors of these databases allocate to developing products in keeping with the characteristics of scientists in these fields. The limited presence of social scientists' and humanists' publications in databases masks much of their scientific output, and lowers the visibility and awareness of their research among members of the scientific community working in other fields.

The geographic scope of social sciences and humanities research is less international than that of the pure, experimental and technical sciences, as a large part of the former research deals with issues of local interest (Nederhof, Luwel & Moed, 2001; Al, Sahiner & Tonta, 2006). This means that the journals used by each group to disseminate their research results differ diametrically: primarily local or national among the former, and much more frequently international among the latter.

Therefore, when social scientists and humanists decide to use scientific papers as a vehicle to convey their information, they usually select national journals, due to the nature of their research and the fact that their chances of being published in international journals are smaller than those of other sciences (López Baena, 2001). Nonetheless, the influence of the local nature of social sciences research on publication habits is waning, for a number of reasons (Hicks, 1999). These include the increased internationalization of national economies and certain technological factors such as the growing use of electronic communication, which enables these scientists to expand their research work to the international arena, and the rising percentage of documents jointly written and published by institutions from different countries (Katz, 1999).

Likewise related to the local character of the research performed by scientists in many fields of the social sciences and humanities is the fact that, unlike the publications of other groups, theirs tend to be in the scientists' native languages (Nederhof, Luwel & Moed, 2001). Garfield (1990) notes that documents published in a language other than English are less visible to the international community; therefore, these research results are disseminated more slowly than findings that are more international in nature and reported in English, the accepted scientific language in many areas of knowledge. This is intensified by the fact that the results of the research conducted by these groups is not usually communicated in scientific jargon, as normally occurs in the pure, experimental and technical sciences; therefore, as they are disseminated in the vernacular, they can reach a larger audience but not necessarily those who specialize in the subject. All of the foregoing factors indicate that national publications are the most suitable channel for the dissemination of this research. This issue has also been studied at Norwegian universities by researchers in close contact with academic circles (Kyvik, 1991). The results were similar to the findings of bibliometric studies conducted in Spain on disciplines such as economics (García Zorita, 2000), psychology (Lascurain Sánchez, 2001) and the humanities (Sanz-Casado *et al.*, 2002).

Each scientific discipline is "expressed" through the channels most appropriate for disseminating the knowledge generated by its researchers. The suitability of the source is related to several factors. One, probably the most important, has to do with the obsolescence of the information being reported. In the case of the humanities and many of the social sciences, the half-life of information is very long, which means that its rate of obsolescence or loss of usefulness is very low. Therefore, monographs are one of the types of documents most widely used by researchers to disseminate the knowledge they generate (Sanz-Casado *et al.*, 2002). Evidently, the low obsolescence rate in these fields of knowledge means that the content of books, whose publication can take several years from the time the research process ends, is current for a long period of time.

However, this situation is very different in the pure, experimental and technical sciences; as the obsolescence rate is much higher, journal articles are the means most widely used by scientists in these fields to disseminate their research results, while monographs are rejected (Hicks, 2004).

All of the foregoing is also very closely related to another characteristic associated with the social sciences and humanities: the pace of the work involved in the research processes is much slower than in other areas. Nevertheless, this has been changing recently in some disciplines, such as archaeology and anthropology, in which the use of methods borrowed from the experimental and technical sciences is stepping up the pace, and thereby providing for speedier scientific results. Therefore, publications in these disciplines have multiplied. As a result, since information becomes obsolete faster, these researchers are using scientific journals to publish their research results.

Furthermore, in certain social sciences, such as economics and psychology, different types of documents (e.g., journals and monographs) may coexist because the intermediate rate of obsolescence of the information they publish makes this information suitable for both types (García Zorita, 2000; Lascurain Sánchez, 2001). This fact is also reflected in Suárez Balseiro's bibliographic review on this subject (2004).

Another factor that influences publication habits in the social sciences and humanities is the pressure on these researchers to have their scientific activity evaluated. For example, in the case of Spain, the scientific system that defines the evaluation criteria on both a national and regional level is beginning to establish a distinction between the criteria applicable to humanists and social scientists as opposed to the rest of the scientific community, taking into consideration how the habits of each group influence its behaviour in acquiring and conveying knowledge.

At this time, scientists are systematically evaluated for different reasons, which may be related to selection or promotion processes in their research careers or to financial incentives linked to their scientific productivity and the quality of their research. Although the situation in the university environment cannot be generally applied to all scientists working in the social sciences and humanities, the most significant research activity in these disciplines takes place in this setting. A study on the scientific production of researchers working in the Autonomous Community of Madrid showed that over 60% of the activity in these areas took place in universities (CINDOC-CSIC, 2004).

In Spain, since the Constitutional Act on Universities (Ley Orgánica de Universidades – LOU, 2001) came into force the scientific activity of these groups is evaluated on a national level by the National Quality Evaluation and Accreditation Agency (Agencia Nacional de Evaluación de la Calidad y Acreditación - ANECA). On a regional level, this evaluation is performed by assessment agencies in each autonomous community. In the Community of Madrid, responsibility for such evaluation is incumbent upon the Quality, Accreditation and Planning Agency (Agencia de Calidad, Acreditación y Prospectiva – ACAP). A review of the evaluation criteria used by these agencies shows that, regardless of a researcher's area of study, one of the common criteria for all is publication in renowned international journals, usually periodicals listed in multidisciplinary databases: Science Citation Index (SCI), Social Sciences Citation Index (SSCI) and Art & Humanities Citation Index (A&HCI), all part of ISI Web of Knowledge. Nonetheless, the possibility that the research submitted by a scientist may not be reflected in such sources is now being taken into consideration. In these cases, other sources to be consulted are specified, although the publications cited are evaluated by the evaluation committee.

The scientific activity of all researchers working out of universities or large public research institutions is evaluated using the same procedures and similar criteria. All must prove that they are in possession of *sexennials*, i.e., merits earned after each six years of service based on outside peer reviews of papers published. This recognition is accompanied by a boost in prestige as well as financial rewards.

However, as noted, due to factors inherent in many social science and humanities disciplines, only a small portion of the output of scientists in these fields, particularly humanities, is reflected in the aforementioned international databases. This situation is the result of both the local nature of the research discussed earlier and the fact that the Web of Science databases provide very limited coverage of publications that are not in the English language. Only these databases and Scopus, however, include the bibliographic references needed to quantify the author or paper citations, through the computerized processing and analysis of these references.

In the case of Spain, the Web of Science's A&HCI database for the humanities offers greater coverage. However, this database does not even provide a citation index, a tool that is very widely used in the evaluation of other disciplines. The lack of an index may be attributed to this group's citation habits, in turn a result of the obsolescence of the literature in this area of knowledge. Indeed, the index that measures the impact of publications (the impact factor) is calculated the same way for the social sciences as for the experimental and technical sciences, i.e., based on the citations a journal receives in the two years following its publication. Consequently, it ceases to be meaningful for the analysis of literature with a much longer, irregular citation period, as in the case of the humanities.

The alternative to such databases, i.e., national resources, generally do not include the bibliographic references to documents, thereby ruling out any citation-related bibliometric analysis of scientific papers. To solve this problem, indices and databases are now being created based on Spanish publications in certain disciplines of the humanities (RESH, IN-RECS, MODERNITAS CITAS, etc [explained below]).

Another of the most unique characteristics of social scientists and humanists, especially in certain fields of the humanities, is researchers' tendency to work alone. The many studies that have stressed this issue (Stone, 1982) have found significant differences between such researchers and experimental and technical scientists, for whom cooperation between authors and institutions is a widespread, well-established practice, which has risen steadily for many years.

The tendency of these scientists to publish individually may be detrimental to their productivity, as several authors found a direct relationship between this productivity and cooperation among several researchers (Endersby, 1996; Durden & Perri, 1995). Nevertheless, even though scientific cooperation is not an easy task, as it requires researchers to communicate and pool their knowledge, there is a crucial factor that motivates them to make the effort: the need to publish (Crase & Rosato, 1992). Researchers in this group may be starting to change their ways, due to the pressure being put on them to strengthen their curricula vitae. However, this trend has not been observed in recent papers published on Spanish researchers in modern history (Fernández Izquierdo *et al*, 2007).

Nevertheless, the new needs arising in the increasingly competitive scientific world, where research ever more urgently requires the contribution of different and complementary expertise from a variety of specialities and disciplines can be expected to influence this pattern of scientific communication and cooperation to bring about a gradual change.

In summary, the characteristics that differentiate scientists depending on their areas of knowledge hinder the creation of bibliometric indicators appropriate to these characteristics. Moreover these indicators are directly related to the content and availability of databases that cover these disciplinary areas. The social sciences and humanities are obviously at disadvantage in this regard, especially in terms of international databases that are clearly biased toward journals in the English-speaking milieu, papers published in English and journal articles, to the detriment of other types of documents and other languages. In this regard, Hicks (1999) also drew attention to the Web of Science's databases, to the effect that "the bibliometric community has adopted the SCI as its de facto standard source [...] However, the more fragmented and polyglot literature of the social sciences is more difficult to cover in a single database".

What can be done to improve scientific evaluation studies in the social sciences and humanities?

Few bibliometric studies have been carried out on the scientific activity of researchers in the areas of the social sciences and humanities, and in many cases, when such studies have been conducted, the research results of these groups have been interpreted in the light of the patterns found in other, previously studied scientific groups. As a result, some very distinctive realities in their scientific work have long been concealed, and these need to be appropriately treated and interpreted in each particular case when the scientific results of researchers in these disciplines are evaluated.

A number of conceptual and methodological considerations that must be taken into account are proposed here to improve the interpretation of the research results of scientists engaging in the social sciences and humanities.

With regard to the conceptual proposals, such scientists have their own deeply ingrained research habits, as already mentioned. These habits have gradually been revealed as different information metrics studies have been undertaken, especially those relating to user and bibliometric studies; (Line, 1971; Brittain, 1979; García-Zorita, 2000; Lascurain-Sánchez, 2001, etc). The difficulty of learning about the scientific activity of these groups in no way speaks of the quality of their research; rather, it is indicative of the strategy followed by the creators and distributors of the major international databases when selecting the information sources they include, and of the fact that these groups have aroused little interest among bibliometric specialists, who have not developed an appropriate conceptual framework to study their production. In this regard, for a little more than a decade, the Information Metric Studies Laboratory (Laboratorio de Estudios Métricos de Información - LEMI), a research group whose aim is to evaluate research and conduct bibliometric and scientometric studies, has undertaken to pursue studies on these disciplines (Sanz Casado *et al.*, 2002; Sanz Casado, Conforti *& collaborators,* 2005, Fernández Izquierdo *et al*, 2007). They have conducted research in Spain on some to obtain theoretical knowledge about the characteristics of these groups in their scientific activity, as well as methodological knowledge intended to develop bibliometric techniques, and particularly indicators, that more closely match their research habits.

Some of the findings of this research have made it possible to identify the need to create benchmarks for comparing the scientific activity of researchers in the same disciplines working out of different institutions or from different countries, to ascertain their progress in the aforementioned activity. The globalization taking place in different segments of society also has a singular effect on the scientific system; thus, the quality of the research conducted by a community, group or individual must be compared to the quality of the research of their national and international peers. To this end, it is essential to have standards that serve as a basis for such comparisons.

As far as the methodological proposals are concerned, the significant boom that is currently taking place in the development of bibliometric techniques that serve as a basis for studies on any scientific group, and especially those groups working in the social sciences and humanities, means that it is essential to address methodological considerations that will provide solutions for many of the issues that have been addressed throughout this paper.

The first of these methodological considerations refers to the sources from which data are obtained. Because of the limited coverage provided by national and international databases, the search for sources where researchers in the social sciences and humanities publish their work poses serious difficulties, and a significant effort must be made to implement strategies to minimize this problem. Some of these strategies depend on finding and developing specific data sources, such as national databases created especially for these groups.

The fact that governments at different levels in Spain have no policy actively aimed at creating their own databases that would include the vast majority of Spanish journals in the areas of the social sciences and humanities, along with the bibliographic references of all the articles published, makes it difficult to systematically undertake bibliometric studies on these groups to ascertain their actual scientific output, as well as the impact and visibility of the research they conduct (Giménez-Toledo, Román-Román & Alcaín-Partearroyo, 2007). The authorities responsible for developing a coherent policy in this area have always used the high cost of producing and maintaining these databases as an excuse; however, with the fast-paced technological advances in today's world, this should not be a problem that would justify the failure to take the action needed to create such data sources.

Research projects are being conducted by scientific groups, aimed at designing and creating specific databases for certain fields in the social sciences and humanities to surmount the problems caused by the lack of a centralized public policy. Even once created, such databases are very difficult to maintain and update, with notable exceptions. These include the IN-RECS project developed at the University of Granada's School of Library and Information Science, which embraces a variety of disciplines in the social sciences. To date, databases have been developed for the following fields: library and information science, economics, education, geography, sociology and psychology (http://ec3.ugr.es/in-recs/). Another initiative is the research project carried out by research groups affiliated with the Institute for Documentary Studies on Science and Technology - Centre for the Humanities and Social Sciences (Instituto de Estudios Documentales sobre Ciencia y Tecnología - Centro de Ciencias Humanas y Sociales - IEDCYT-CCHS), known as RESH (Spanish Social Sciences and Humanities Journals) (http://resh.cindoc.csic.es).

Finally, another project with similar characteristics merits mention. This project one is being conducted by the Information Metric Studies Laboratory (LEMI) of Carlos III University of Madrid, the History Institute and IEDCYT-CCHS, both attached to the CSIC (Spanish Council for Scientific Research - Consejo Superior de Investigaciones Científicas). This project has spawned the citation index known as Modernitas Citas (www.moderna1.ih.csic.es/emc), which is being developed with data taken from modern history journals. As for RESH, the first stage of this project consisted of selecting the journals in this field with the highest quality. To this end, a variety of evaluation criteria were used, ranging from quantitative to qualitative. The bibliographic references for each article have been included in this index, to create a tool with which to both establish queries by source publication and retrieve the works cited.

The three projects discussed above take account of each journal's citation index, whereby the use and influence of each in its scientific field can be assessed. As they analyze the citations received by journals, these projects can be used to ascertain the visibility of Spanish publications in these disciplines. Similarly to the Web of Science's databases, they can also be classified according to this visibility, and their evolution over time can be monitored.

The reports of universities and other research centres are another important source of data to be taken into consideration in studies evaluating these scientific groups. Often, as a result of the national database related problems discussed earlier, and of the fact that universities and other research centres need to periodically and exhaustively analyze the scientific production of their teaching staffs to evaluate progress in these areas, programmes designed to reflect all of the scientific activities conducted by the teaching staff on a yearly basis are underway in academic domains.

These reports are an extraordinarily precise source of information for evaluating the research conducted by teaching staff, to implement a policy for the control and distribution of available resources that is fair to the members of the institution and appropriate to each individual's efforts in his or her scientific work. They would also be one of the best data sources for the study of the research output of social scientists and humanists, as such reports contain very detailed information about these groups' scientific activity, information that is very difficult to find for the reasons discussed earlier.

The researchers' curricula vitae are another source that should be taken into consideration; they are particularly useful for bibliometric studies that involve the individualized analysis of research activity. The fact that many regional and national agencies are evaluating the scientific production of researchers makes it necessary, indeed, essential, to systematically ensure that all researchers' curricula vitae are complete, and in particular those of scientists who work in the areas of the social sciences and humanities, for their activity is more difficult to ascertain. In order to standardize this task, several Spanish universities have begun to develop a new tool to manage research activity, known as Universitas XXI. This tool represents a significant advance, as its structure includes a large number of fields that have been adapted to the specific case of scientific activity in a university environment, in order to gather information about most of the contributions made by researchers and especially those in the social sciences and humanities.

Other important methodological aspects to be taken into consideration when studying the scientific output of researchers in the social sciences and humanities are data acquisition, processing and analysis. Specific methodologies should be developed to adapt these processes to their research habits and characteristics.

In the acquisition of scientific production data, longer time frames must be taken into consideration, and subject searches must be more exhaustive. The reason is that the research times of groups working in these disciplines are usually longer than those of the experimental and technical sciences, and the types of documents they publish tend to be monographs. This is why it is difficult to obtain accurate insight into the scientific activity of these researchers when short time periods, such as normally taken into consideration for other disciplines, are used.

Subject searches, in turn, must be exhaustive because of the difficulty of obtaining the scientific production of these researchers when querying specialized databases, for their research usually covers a wide range of subjects that are not as strictly delimited as the research of scientists in experimental and technical sciences.

Another factor to take into consideration with regard to the acquisition of data is the type of publication in which these researchers disseminate their work; as these types can vary widely and are usually less visible, data may be more difficult to find, as discussed earlier.

As far as data analysis is concerned, the fact that researchers in each group present results that are typical of the discipline in which they work must be taken into consideration; therefore, they must be analyzed and conclusions must be reached in the context of this discipline and not extrapolated to other fields (Kyvik, 1991).

Finally, special emphasis should be placed on obtaining bibliometric indicators adapted to the characteristics of these researchers. This is another essential aspect of studies evaluating the scientific activity of researchers in the social sciences and humanities. With the use of these tools, the evaluation studies that can be conducted on these groups are more complete, yielding a more accurate reflection of their scientific situation. Specific indicators are being obtained for this purpose, which in some cases differ from and in others supplement the indicators used in bibliometric studies conducted in the experimental and technical sciences. They are essential for acquiring data from different perspectives relating to the particular characteristics of their activity.

Bibliometric studies of these groups must also obtain a significant number of indicators to reflect the variability of their research, and seek the convergence of these indicators, to reveal the peculiarities of their work from different perspectives.

A great effort is being made at this time to develop multidimensional or relational bibliometric indicators, which are the most appropriate indexes for analyzing scientific activities as heterogeneous as those of researchers in the social sciences and humanities. These indicators are yielding a holistic view of scientific activity through the simultaneous comparison of different characteristics involved in this activity.

References

Al, U.; Sahiner, M. & Tonta, Y. (2006) Arts and Humanities Literature: Bibliometric Characteristics of Contributions by Turkish Authors. *Journal of the American Society for Information Science and Technology*. Vol. 57, nº 8, p. 1011-22.

Brittain, J. M. (1979) *Information and its uses. A Review with special reference to the Social Science*. Bath: Bath University Library.

CINDOC (CSIC). (2004) *Indicadores de Producción Científica y Tecnológica de la Comunidad de Madrid (PIPCYT): 1997-2001*. Madrid: Comunidad de Madrid.

Crase, D. & Rosato, F.D. (1992) Single versus multiple Authorship in professional journals. *JOPERD*: 28-31.

Durden, G. C. & Perri, T. J. (1995) Coauthorship and Publication Efficiency. *Atlantic Economic Journal*. Vol, 23, nº 1, p. 69-76.

Endersby, J. W. (1996) Collaborative Research in the Social Sciences: Multiple Authorship and Publication Credit. *Social Science Quarterly*. Vol. 77, nº 2, p. 375-92.

España. *Ley Orgánica, de 21 de diciembre de 2001, de Universidades*, Boletín Oficial del Estado, 24 de diciembre de 2001. 309: 49400-49425

Fernandez Izquierdo. F.; Román Román, A.; Rubio Liniers, C.; Moreno-Díaz del Campo, F.J.; Martín Moreno, C.; García Zorita, C.; Lascurain Sánchez, M.L.; Efraín-García, P.; Povedano, E. & Sanz Casado, E. (2007) Bibliometric study of Early Modern History in Spain based on bibliographic references in national scientific journals and conference proceedings. In Torres Salinas, D. & Moed, H.F. *Proceedings of ISSI 2007. 11th International Conference of International Society for Scientometrics and Informetrics*. Madrid: CSIC, p. 266-71.

García Zorita, J. C. (2000) *La actividad científica de los economistas españoles, en función del ámbito nacional o internacional de sus publicaciones: estudio comparativo basado en un análisis bibliométrico durante el período 1986-1995* [Doctoral thesis]. Getafe: Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.

Garfield, E. (1980) Is information retrieval in the arts and humanities inherently different from that in science? The effect that ISI's citation index for the arts and humanities is expected to have on future scholarship. *Library Quarterly*. 50: 40-57.

Garfield,, E. (1990) The languages of science revisited: English (only) spoken here? *Current Contents*, nº 31, p. 3-18.

Giménez-Toledo, E., Román-Román, A. & Alcaín-Partearroyo, M.D. (2007). From experimentation to coordination in the evaluation of Spanish scientific journals in the humanities and social sciences. *Research Evaluation*. Vol. 16, nº2, p. 137-48.

Hicks, D. (1999) The difficulty of achieving full coverage on international social science literature and bibliometric consequences. *Scientometrics*. Vol. 44, nº 2, p. 193-215.

Hicks, D. (2004) The four literatures of Social Sciences. En: Moed, H. F.; Glänzel, W. y Schmoch, U. (Eds). *Handbook of Quantitative Science and Technology Research*. Dordrecht: Kluwer Academic Publishers; 473-496.

Katz, J. S. (1999) *Bibliometric indicators and the social sciences*. Brighton: SPRU, University or Sussex.

Kyvik, S. (1991) *Productivity in Academia*. Oslo: Norwegian University Press.

Lascurain Sánchez, M. L. (2001) *Análisis de la actividad científica y del consumo de información de los psicólogos españoles del ámbito universitario durante el período 1986-1995* [Doctoral thesis]. Getafe: Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.

Line, M. B. (1971) The information uses and needs of social scientists: An overview of INFROSS. *ASLIB Proceedings*. Vol. 23, nº 8 p.: 412-34.

López Baena, A. J. (2001) *Innovaciones en la evaluación y mejora de la investigación científica: una perspectiva institucional* [Doctoral thesis]. Córdoba: Unidad para la Calidad de las Universidades Andaluzas.

Nederhof, A. J.; Luwel, M. & Moed, H. F. (2001) Assessing the quality of scholarly journals in Linguistics: An alternative to citation-based journal impact factors. *Scientometrics*. Vol. 51, nº 1, p. 241-65.

Sanz Casado, E.; Conforti, N., & collaborators. (2005) Análisis de la actividad científica de la Facultad de Humanidades de la Universidad de Mar de Plata, durante el período 1998-2001. *Revista Española de Documentación Científica*. Vol. 28, nº 2, p. 196-205.

Sanz, E.; Castro, F.; Povedano, E.; Hernández, A.; Martín, C.; Morillo-Velarde, J.; García-Zorita, C.; Nuez, J. L. de la & Fuentes, M. J. (2002) Creación de un índice de citas de revistas españolas de Humanidades para el estudio de la actividad investigadora de los científicos de estas disciplinas. *Revista Española de Documentación Científica*. Vol. 25, nº 4, p. 443-54.

Siatri, R. (1999) The evolution of user studies. *Libri*, Vol.49, nº 3, p. 132-41

Stone, S. (1982) Humanities scholars: information needs and uses. *Journal of Documentation*. Vol. 38, nº 4, p. 292-313.

Suárez Balseiro, C. A. (2004) *Perfiles de actividad científica de los departamentos de la Universidad Carlos III de Madrid: un estudio con variables de recursos y resultados del proceso científico durante el período de 1998 a 2001* [Doctoral thesis]. Getafe: Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.

Addresses of congratulating authors:

ISABEL IRIBARREN-MAESTRO
Carlos III University of Madrid, C/Madrid, 126. Getafe 28903 Madrid (Spain)
Email: iiribarr@bib.uc3m.es

MARÍA LUISA LASCURAIN-SÁNCHEZ
Carlos III University of Madrid, C/Madrid, 126. Getafe 28903 Madrid (Spain)
Email: mlascura@bib.uc3m.es

ELIAS SANZ-CASADO
Carlos III University of Madrid, C/Madrid, 126. Getafe 28903 Madrid (Spain)
Email: elias@bib.uc3m.es

Persson's universe of bibliometrics – Has his mapping changed the discipline?

Martin Meyer[1] and Wolfgang Glänzel[2]

[1] SPRU Science and Technology Policy Research, Freeman Centre, Brighton, (UK)
[2] Centre for R&D Monitoring (ECOOM), Katholieke Universiteit Leuven, Dekenstraat 2, (Belgium)

Introduction

Nowadays we are faced with a plethora of impressive maps reflecting the structure of the research landscape and the universe of documented scholarly communication, expressing important positional and relational aspects by measuring the distance among and similarity of individuals objects and clusters. Beyond doubt, this cartography of science and technology permits insight into important aspects of the cognitive structure of scientific research, helps monitor the evolution, that is, the emergence, convergence and decline of research topics and disciplines and thus the changing universe of science.

Studying these maps and their frequently appearing updates, the question arises of how scientists, librarians and information practitioners can translate their observation into the practical needs of their daily work. In other words, how can scientific information provided by these maps broaden the consciousness of what is relevant for the own and the colleagues' research and thus possibly improve the efficiency of communication in science as well? The co-citation based *Atlas of Science* developed and issued by the Institute for Scientific Information (ISI) was actually one of the first endeavours in mapping the cognitive structure of the research landscape and this atlas was considered a new kind of 'review literature' which might also be suited to help students in the choice of career in science (Garfield, 1975, 1988). This also implies that (future) scientists might learn from these visualisations, in particular, what is 'useful' and what might be 'hot' in their discipline, and might thus be able to better find and position their own research tasks. This function of science maps reaches far beyond the scope of information science, in general, and bibliometrics, in particular.

A second question arises from this perspective, namely: could this effect be strengthened if scientists can prepare their own maps to better understand their own role and position in the network of scientific communication? The answer is given by Olle Persson's work; it is not a suddenly formulated clear and unique answer but – as we will see in the following – the solution was found in presenting a toolbox and in continuous interaction with its users. This solution can be considered an extension of Persson's notion of online bibliometrics as a research tool for everybody (Persson, 1986). In order to "measure" this effect we will compare utilisation and impact of this toolbox in the scientific community with the impact of Persson's "regular" scientific work.

Bibmap and Bibexcel – a Swedish look at science

Most results of the cartography of science are presented as results of scientific research in information science or as commercial products as they are provided. Although some of these products are "tailor-made", a real interaction between producers and users of the maps remains rather the exception to the rule. Of course, there is public-domain software for plotting the results of some analysis of data downloaded from bibliographic databases or for conducting network analysis based on such data (e.g., Pajek, see Batagelj and Mrvar, 2002). However, the application of such tools requires careful data processing prior to the use. Their application was therefore rather restricted to experts or at least to researchers with basic skills in both bibliometric techniques and computer programming. The gap between the immediate download from a citation index and a user-friendly and completely customisable tool to be used for analysis and visualisation of bibliometric data was bridged by Olle Persson's ingenious design of *Bibexcel* as early as the beginning 1990s. The idea is intimately connected with 'BIBMAP' (Persson et al., 1992). The basic idea behind Bibexcel is to prepare and process the downloaded records so that those can be used as the direct input of any advanced spreadsheet application. The author's description of this product sounds rather modest and sober (Persson, 2009).

> *"Bibexcel is designed to assist a user in analysing bibliographic data, or any data of a textual nature formatted in a similar manner. The idea is to generate data files that can be imported to Excel, or any program that takes tabbed data records, for further processing."*

Nevertheless, this tool has revolutionised the work of many users word-wide. Its use had been documented through citations or references to its author and/or website. Traditional abstract and citation databases are therefore not suitable for retrieving the application of Bibexcel. Hence,we decided to use the web and, above all, *Google Scholar* (Google Inc., Mountain View, CA, USA). In order to measure the impact of Bibexcel on the community, we compare the outcomes with citations received by other Persson's work (excluding Bibmap and Bibexcel). The reference data have been retrieved in and extracted from the *Web of science* (WoS, Scientific – Thomson Reuters, Philadelphia, PA, USA). Citing papers have been retrieved for the same period as for the documents in Google Scholar (from 1992 on). Besides Persson's papers indexed in the WoS since 1992, his two contributions in the handbooks of Quantitative Science and Technology Research (Ed: AFJ van Raan, Elsevier, 1988; Eds: HF Moed, W

Glänzel, U Schmoch, Kluwer, 2004) have been taken into account. The results are analysed and discussed in the following section.

The impact of Bibexcel on the community

The direct comparison between the utilization of Bibexcel and the reception of Persson's research work by the community immediately reveals interesting divergences in the use of information. In this context, we have to stress that the validity of the interpretation of the results taken from different sources will be somewhat limited. This has to do with different quality criteria of indexing documents in the two databases and the "language barrier". While almost all citing papers in the WoS are in English, a great deal of documents found in Google Scholar are in other languages (Swedish, Spanish, Chinese etc.). Nevertheless, we are confident that the outcomes of this comparison are sufficiently significant. Above all, the question of who are the Bibexcel users and what they are dealing with will be in the foreground. Before we tackle this issue, will have a view at Bibexcel utilization in the mirror of Google Scholar (see Table 1).

*Table 1 The use of Bibmap/Bibexcel in the mirror of Google Scholar (*Citations counts estimated for the complete year)*

PY	Citations received	
	Increments	Cumulated
1992	2	2
1994	1	3
1996	1	4
1998	1	5
1999	2	7
2000	4	11
2001	6	17
2002	9	26
2003	8	34
2004	14	48
2005	20	68
2006	15	83
2007	23	106
*2008	18	124

Since data were retrieved in 2008, we had to extrapolate on the basis of the trend of the previous ten years. For the period January–July 2008 we have counted 10 citations. Based on the exponential model suggested by Figure 1, we have applied an exponential extrapolation to estimate the complete citation impact for 2008.

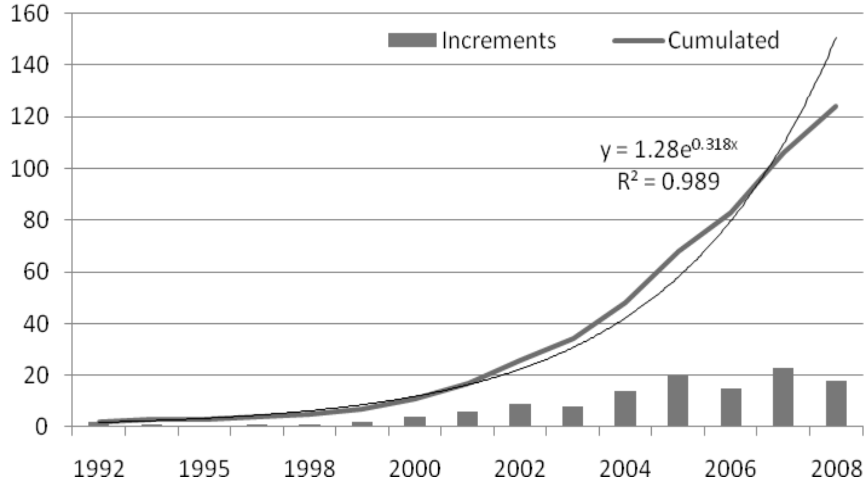

Figure 1 Growth of Bibexcel application as measured by citations recorded in Google Scholar

The geographical distribution of Bibexcel users immediately shows a considerable deviation from that of citers of Olle Persson's research work (cf. Table 2). The fact that the software is mostly used in Sweden is obvious since Persson as the author and inventor of this tool is assumed to used it frequently in his daily work. The rest of the distribution substantiates that Bibexcel is "a research tool for every man" indeed (cf. Persson, 1986). Scientists from less developed and newly industrialised countries, above all, in Latin America, Africa and Asia are well-represented. By contrast, we find authors from those countries on top of the ranking list of Persson citers, which are also otherwise among the most active in bibliometric research.

Table 2 The geographical distribution of users of Bibexcel in comparison with that of citers of Persson's research work (share ≥2.5%)

Rank	Bibexcel			Persson		
	Country	Cites	Share(%)	Country	Cites	Share(%)
1	Sweden	21	18.1	USA	65	16.0
2	England	14	12.1	Netherlands	44	10.9
3	Denmark	13	11.2	Sweden	43	10.6
4	USA	10	8.6	England	40	9.9
5	Spain	9	7.8	Spain	33	8.1
6	China	8	6.9	Canada	30	7.4

Rank	Bibexcel			Persson		
	Country	Cites	Share(%)	Country	Cites	Share(%)
7	Cuba	8	6.9	Germany	28	6.9
8	Finland	6	5.2	Belgium	24	5.9
9	India	6	5.2	France	21	5.2
10	Mexico	6	5.2	Denmark	19	4.7
11	South Africa	6	5.2	Japan	18	4.4
12	Belgium	4	3.4	China	18	4.4
13	Germany	4	3.4	India	15	3.7
14	Argentina	3	2.6	Finland	14	3.5
15				Australia	13	3.2
16				Hungary	11	2.7

The second question concerns the topic or context in which Bibexcel is used. The large share of users in information science in both populations is, of course, not surprising. Nevertheless, there is an essential deviation outside this user group (see Tables 3 and 4). While more than 50% of the papers referring to or reporting utilisation of Bibexcel can be assigned to subjects outside information science and a significant user group could be found even in the life sciences, citers of Persson's research work are according to the ISI database rather restricted to "main discipline" of information science. The disciplines in Computer science are "greyed out" because those are a by-effect of the multiple assignment of most information-sciences journals (JASIST, Scientometrics, JOI, Journal of Information Science, IP&M, etc.). Only the weight field of Operations and Management science and related fields roughly coincides in both populations. In this context we have to mention that we avoided multiple assignments in the Bibexcel user group while we just used the not uniquely defined ISI Subject categories for the papers citing Persson's research work.

Table 3 Domains in which Bibexcel is used (based on the share of citing documents indexed in Google Scholar; ≥5%)

Field	**Share (%)**
Information Science & Scientometrics	46.7
Health & Medicine	17.1
Operations & Management science	9.5
Social sciences & Education	5.7
Others	21. 0

Table 4 ISI Subject Categories in which Persson is cited (based on the share of citing documents indexed in the Web of Science; ≥5%)

Field	Share (%)
Information Science & Library Science	68.1
Computer Science, Interdisciplinary applications	34.8
Computer Science, Information systems	20.2
Management	9.1
Planning & Development	6.2

Now it is time to have a look at the impact of Bibexcel from the dynamic perspective. The annual change of documented Bibexcel utilisation was already presented in Table 1. Figure 1 presents both annual increments and cumulated number of citations in Google Scholar. The trend is estimated on the basis of an exponential regression. This model provided the best fit with a strong correlation with a correlation coefficient of $r = 0.995$. This picture is contrasted by the evolution of citations to Persson's research work which, in turn, can be characterised as a sub-exponential but supra-linear growth (see Figure 2). The power model proved to provide the best estimate. The correlation coefficient amounts to $r = 0.998$ and the regression equation is an almost perfect quadratic function ($y = x^2$). Both cases reflect a supra-linear growth of the impact of Persson's work, where the popularity of Bibexcel outruns the effect of the research impact by approaching a quasi-exponential growth although this trend seems to somewhat drop in the last two year (cf. Figure 1).

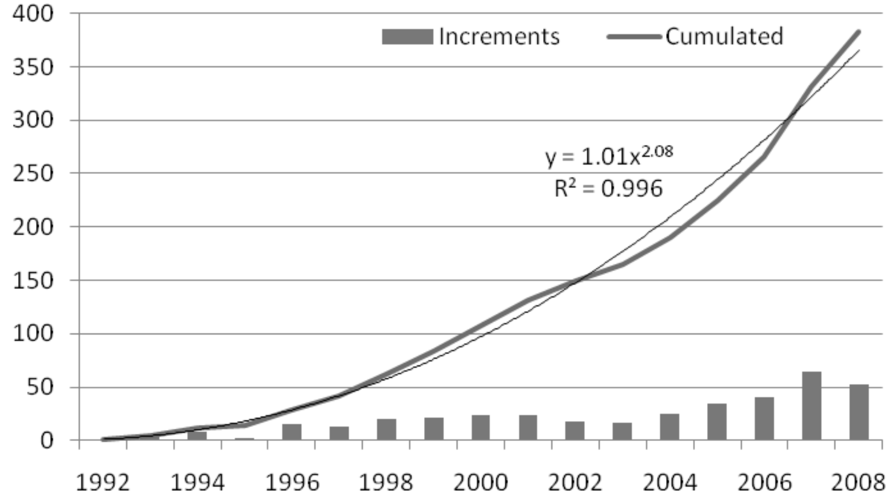

Figure 2 Evolution of citations received by Persson's research papers as reported by the Web of Science

Another way of looking at the citation of Bibexcel is to think of its adoption by the community as the diffusion of an innovation. There are several approaches modelling these diffusion processes (e.g. Rogers, 1995; Bass, 1969), most of which suggest an s-shaped curve to describe the (accumulated) uptake of an invention. Rogers (41995) made a distinguished different categories of adopters – innovators (accounting for about the first 2.5% of new users), early adopters (13.5%), early majority (34%), late majority (34%), and laggards (16%). Roughly, the growth in the application of Bibexcel compares to the early adopter stage in Rogers' model. Assuming that the model applies in this instance, this would suggest that there is still plenty of room for potential new users to adopt Bibexcel. A closer inspection of the citation data in *GoogleScholar* suggest that since the early years of our decade (especially from 2003 onwards), Bibexcel has been increasingly used outside the Library and Information Science community, winning over new users in fields, such as health, education, sociology as well as technology and engineering management.

Conclusion

From measuring and comparing the reception and utilisation of Olle Persson's research work and his software tool *Bibexcel* we may conclude that Bibexcel proved indeed a tool for everybody in research and application not only in information-science related disciplines. This tool is widely used within as well as outside the main areas (both geographically and in a cognitive sense). The increasing popularity of Bibexcel can be characterised as even being exponential. Persson's tool attracts more and more new users in other communities and is becoming more and more what DeSolla Price (1984) once called an 'instrumentality' – an instrument, technique, or procedure that serves as a driver of research across fields and disciplines in science and technology.

References

Batagelj V, Mrvar A (2002). Pajek – analysis and visualization of large networks. Graph Drawing, 2265, 477–478.

Bass FM (1969), A new product growth model for consumer durables. Management Science, 15, 215-227.

Garfield E (1988), The encyclopedic ISI Atlas of Science launches three new sections: Biochemistry, Immunology, and Animal & Plant Science, Current Contents, (7), 3-8.

Garfield E (1975), ISI's Atlas of Science may help students in choice of career in science, Current contents, (29), 5-8.

Persson O (1986), Online Bibliometrics. A research tool for every man. Scientometrics, 10, 1-2, 69-75.

Persson O, Stern P, Holmberg K-G (1992), BIBMAP: a toolbox for mapping the structure of scientific literature. In: Representations of Science and Technology. Leiden, the Netherlands, DSWO Press, p. 189-199.
Persson O (0000), BIBEXCEL. Accessible via: http://www8.umu.se/inforsk/Bibexcel/index.html.
Price, DJdS (1984). The science/technology relationship, the craft of experimental science, and policy for the improvement of high technology innovation. Research Policy 13(1), 3-20.
Rogers EM (1995), Diffusion of innovations. 4th edition, Free Press: New York.

Addresses of congratulating authors:

MARTIN MEYER
SPRU Science and Technology Policy Research, Freeman Centre, Brighton, UK)
E-mail: M.S.Meyer@sussex.ac.uk

WOLFGANG GLÄNZEL
Centre for R&D Monitoring (ECOOM), Katholieke Universiteit Leuven,
Dekenstraat 2 (Belgium)
E-mail: Wolfgang.Glanzel@econ.kuleuven.be

The most influential editorials

Ronald Rousseau

KHBO, Industrial Sciences and Technology, Zeedijk 101, 8400 Oostende (Belgium)
K.U.Leuven, Department of Mathematics, Celestijnenlaan 200B, 2001 Leuven (Heverlee), (Belgium)

Abstract

This article studies the citation influence of editorials. It is found that only a few journals publish highly cited editorials (*New England Journal of Medicine, Nature* and *Science*) and that their authors come overwhelmingly from the USA, with Harvard as the leading university. It turned out that this article is much more about methodology than about finding the most influential editorials.

Introduction

An editorial is like a short essay or a written speech. Most speeches make no or little impression and are forgotten the moment they are spoken. Yet, some have a lasting influence: take for instance Dr. Martin Luther King Jr's "I have a dream" speech, Mahatma Gandhi's "Quit India" speech and J.F. Kennedy's "Ich bin ein Berliner" speech.

Editorials are often just a presentation of the contents of a journal's issue including maybe some comments on salient points. Other editorials contain reflections related to some special event. Yet, every now and then an editorial contains a forceful message, a call for action, perhaps a little known scientific fact with far-reaching consequences is brought into the limelight. These are the editorials that are remembered by fellow scientists. Whatever the concrete contents of such editorials, whatever the concrete journal and field in which they are published, one may rightly say that the editorial itself is not a scientific contribution in that particular field. It is a literary piece, to be classified as an article in the humanities, even if written by a professional physicist, cell biologist or economist. But, are these the editorials that are most cited? In this article, dedicated to Olle Persson on the occasion of his 60th birthday, we investigate which editorials are highly cited and in which journals they are published.

Definition

Definition

How to define an editorial? An editorial in a regular newspaper is usually a short article expressing an opinion or point of view, written by the main editor or another member of the publication staff. Yet, as mentioned above, in scientific articles an editorial is somewhat different. As it is, moreover, impossible to read all articles that might be considered to be an editorial, we used the Web of Science's concept of 'editorial material' as a starting point. After some try-outs we opted for the following definition: an editorial is a publication classified as 'editorial material' (in the WoS), of length at most 3 pages, with a reference list of at most 10 items and written by one person (or published anonymously). We recall that Garfield (1987) has published an internal (ISI) grading algorithm for determining so-called substantial articles that are not published in the same way as regular articles or reviews. These substantial articles are published as letters, editorial materials, comments but are actually research articles and should be treated as such. The editorials we are looking for are not substantial research articles. If Thomson Reuters still uses an algorithm similar to the one published by Garfield, then these substantial articles are automatically eliminated from the category of editorial materials. We think, however, that this is not the case, as we found many 'substantive' articles among the so-called editorial material. So, for this reason we had to delineate the set of editorials further.

Data Collection

For the period 1975 – 2008 we collected each year's five most-cited editorials (according to the definition above). This leads to a total of 172 articles (two ties on the fifth place). Data collection took place during the first week of February 2009. When checking some results we found out that some articles registered in the WoS as a single-authored paper were actually written by a committee, using a group name; or were just misrepresented in the WoS. The following publication is a case in point:

The article

> A working formulation for the standardization of nomenclature in the diagnosis of heart and lung rejection: Heart Rejection Study Group.
> By The International Society for Heart Transplantation: Billingham, ME, Cary NR, Hammond ME, Kemnitz J., Marboe C., McAllister HA, Snovar DC, Winters GL, Zerbe A.
> *Journal of Heart Transplantation* (1990), vol.9, pp. 587-93

is registered in the Web of Science (with zero citations), but all citations are assigned to:

ME. Billingham
International Society for Heart-Transplantation
Journal of Heart Transplantation (1990), vol.9, p. 587-587

This article received 1,241 citations (February 2009), hence was first included in our list of most-cited editorials. We checked this case and found that on page 587 Margaret Billingham, the first author of the actual article wrote a short introductory note (one that rightly could be called an editorial). However, ISI, assigned all citations to the standardization of nomenclature article to this short introductory note. Actually, the journal itself is not very clear, but its table of contents is. Indeed: this journal issue begins with three sections: *President's Message* (this is Margaret E. Billingham's short editorial), *International Society for Heart Transplantation* and *Original Articles*. Clearly, ISI has used the name of the section as the title of Billingham's untitled editorial. The section *International Society for Heart Transplantation* contains two proposals for standardization: one by the Heart Rejection Study Group and one by the Lung Rejection Study Group (beginning on page 593 and cited 343 times). This example highlights one of the methodological problems related to the study of scientific editorials.

Results

The Top 10

We begin by showing the top 10 most-cited editorials published over the period 1975 – 2008. Three of these articles are from Harvard University, two of which written by Nobel laureate Walter Gilbert. The first one has one author, who, however, acts for a committee.

Table 1 The 10 most-cited editorials (according to our definition)

Rank	Document	No. citation
1	Manfred Zimmermann Ethical guidelines for investigations of experimental pain in conscious animals Pain (1983), vol.16, pp. 109-110; 9 references Published as: Guest editorial	2452
2	Walter Gilbert Why genes in pieces? Nature (1978), vol. 271, p.501; 10 references Published as: News and Views	1520

Rank	Document	No. citation
3	Lewis C. Cantley The phosphoinositide 3-kinase pathway Science (2002), vol.296, pp. 1655-1657; 8 references Published as: Viewpoint	1257
4	Craig L. Hill Introduction: Polyoxometalates - Multicomponent molecular vehicles to probe fundamental issues and practical problems. Chemical Reviews, (1998), vol.98, pp. 1-2; 0 references Published as: Introduction	761
5	Walter Gilbert Origin of life – the RNA world. Nature (1986), vol. 319, p.618; 9 references Published as: News and Views	711
6	Cyrus Chothia Proteins – 1000 families for the molecular biologist Nature (1992), vol.357, p.543-544; 10 references Published as: New and Views	543
7	David Wynford-Thomas P53 in tumor pathology – can we trust immunocytochemistry? Journal of Pathology (1992), vol.166, 329-330; 6 references Published as: Editorial	446
8	Robert J. Bodnar Revised equation and table for determining the freezing-point depression of H_2O-NaCl solutions Geochimica et Cosmochimica Acta, (1993), vol.57, pp.683-684; 2 references Published as: Scientific comment	413
9	Arnold S. Relman Assessment and accountability – the third revolution in medical care New England Journal of Medicine (1988), vol.319, pp.1220-1222; 4 references Published as: Editorial	375
10	Christopher S. Foote Definition of type I and type II photosensitized oxidation Photochemistry and Photobiology (1991), vol.54, p. 659; 7 references Published as: Guest Editorial	370

Journals- Fields

In which journals are these most-cited editorials published? Essentially there are three journals in which highly-cited editorials are published: the *New England Journal of Medicine* occurs 39 times in the list, *Nature* 30 times and *Science* 24 times. The next journal in the list is *Chemical Reviews* with 5 editorials. Yet, it is no surprise that the size-frequency list of journals follows a Lotka (power law) distribution: $f(y) = \frac{0.73}{y^{2.43}}$, where f(y) denotes the relative number of journals with y contributions to this list. Fitting has been performed using the LOTKA program (Rousseau & Rousseau, 2000).

Next we determined to which fields these journals belong. Writing successful or at least highly cited editorials is a two field business: *Multidisciplinary Sciences* (54) and *Medicine, General & Internal* (51). Also here the complete list of fields follows a Lotka distribution: $f(y) = \frac{0.665}{y^{2.18}}$, where now f(y) denotes the relative number of fields with y contributions to this list.

Addresses

In which countries do the writers of highly cited editorials work? Several of these editorials do not have an address, but among those who have the USA leads by a wide margin. Table 2 shows this elite group of countries. The UK figure consists of 18 contributions from England, 2 from Wales and 1 from Scotland. Even this short list has Lotka characteristics ($f(y) = \frac{0.43}{y^{1.58}}$).

Table 2 *Countries*

Country	Number of contributions	Country	Number of contributions
USA	93	Switzerland	2
UK	21	Australia	1
Germany	9	Canada	1
Japan	4	Italy	1
The Netherlands	3	Sweden	1
Denmark	2		

As to universities or institutes: Harvard has 10 contributing articles, the National Institute of Health (NIH) 5 as does the University of California, Berkeley; the University of Texas has 4 contributions and so does Oxford University. One could say that this is a mini-list of top universities and institutes.

Some further remarks

The most-cited editorials have, by the definition we used, between zero and ten references. One would expect that a 'real' editorial has no references. This turned out to be true, in the sense that zero references is indeed the mode of the number of references. Yet, many so-called editorials have a large number of references (7 to 10). Details are shown in Table 3.

Table 3 Number of references

Number of references	Number of articles	Number of references	Number of articles
0	37	6	11
1	9	7	21
2	6	8	12
3	6	9	18
4	9	10	31
5	12		

We also performed a search to find the most-cited editorial in the information sciences. Several articles by Eugene Garfield almost made it, but their references lists were too long. The same is true for Jean Tague-Sutcliffe's introduction to a special issue on informetrics in Information Processing and Management (Tague-Sutcliffe, 1990). The one we found is the late Michael Moravcsik's announcement in Scientometrics of the first Derek de Solla Price awardee:

M.J. Moravcsik – 8 citations
Address at the presentation of the 1st Price, Derek D. award to
Garfield, Eugene on December 20, 1984
Scientometrics (1985), vol.7, pp. 143-144; no references

Conclusions

Newspaper editorials have been studied in Journalism and Communication Science, and also Garfield has paid attention to the interesting ideas that are published in editorials (Garfield, 2000), but to the best of our knowledge this is the first article trying to highlight influential editorials in academic journals.

This investigation turned out to be a pilot study for those who want to use editorials as part of a larger research evaluation exercise, or as a part of an investigation on the structure of science. Two serious problems have been detected. The first is related to the definition of an editorial. Probably, the definition we used is still too broad. Several of the 'editorials' we found can better be described as short scientific communications. We, most certainly, did not find an equivalent to the famous speeches mentioned in the introduction. Taking into account that the first characteristic we applied to define an editorial was to be considered 'editorial material' in the Web of Science, this shows how this category is actually a very mixed bag.

The second problem is related to the indexing of the Web of Science. We cannot be sure if the single-authored articles (based on the records in the WoS), we retained as editorials are actually single-authored.

Clearly, these problems can only be solved by using a more refined definition of 'an editorial' and by having access to the original articles, namely the supposed editorials themselves, in order to visually check their exact content. Maybe interesting proposals, or calls to arms (similar to the famous speeches) can also, or even more often, be found as *Letters to the editor* or *Correspondences*. One final idea: can some types of editorials be used to track new or even future developments in science? If the answer is yes, this would make them a data mining tool, among many other ones, to predict new and emerging trends.

Acknowledgments

I thank Nadya Verschelden and Kurt Noppe (Ghent University, Belgium) for their help in finding and describing Margaret Billingham's article, and Mikael Graffner (Lund Univ.) for checking Zimmermann's contribution. Raf Guns (Antwerp University) is acknowledged for useful comments on an earlier version.

References

Garfield, E. (1987). Why are the impacts of the leading medical journals so similar and yet so different? Item-by-item audits reveal a diversity of editorial material. Current Comments, #2, p.3 January 12.

Garfield, E. (2000). Foreign language editorials should be translated for the Web. The Scientist, 14(9), p. 6.

Rousseau, B. & Rousseau, R. (2000). LOTKA: a program to fit a power law distribution to observed frequency. Cybermetrics, 4(1), paper 4. Data Retrieved May 8, 2009 from: http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html

Tague-Sutcliffe, J. (1992). An introduction to informetrics. Information Processing and Management, 28, 1-3.

Address of congratulating author:

Ronald Rousseau
KHBO, Industrial Sciences and Technology, Zeedijk 101, 8400 Oostende (Belgium)
K.U.Leuven, Department of Mathematics, Celestijnenlaan 200B, 2001 Leuven (Heverlee), (Belgium)
Email: ronald.rousseau@khbo.be

Publication patterns in all fields

Gunnar Sivertsen

NIFU STEP, Wergelandsveien 7, N-0167 Oslo (Norway)

My first encouter with bibliometrics was in April 1988 in Elsinore, Denmark at a conference about scholarly publishing that was organized by the Nordic Council of Ministers. I remember well how bibliometrics suddenly personalized as the floor was taken by a bearded Swede in tweed and jeans. He demonstrated empirically the need for further internationalization of the social sciences in Scandinavia by using articles in scientific journals as a data source (Persson 1988). It was amazing.

I attended the conference because I had been invited to speak about international marketing of scholarly publications (Sivertsen 1988). At that time, I was the editorial director of the journals department at Scandinavian University Press. I knew nothing about bibliometrics, although I do recall that my office – for a reason I could not understand – subscribed to an obscure journal from Eastern Europe with an orange cover and a futuristic title. Anyways, I never read nor opened the journal before Olle demonstrated what bibliometrics could do in April 1988.

Only a few weeks later, I shifted over to my present position as a researcher at NIFU STEP and started with bibliometrics from day one. It did not take long before Nordic collaboration became both a subject matter and a research practice resulting in co-publications in Norwegian (Luukkonen, Persson & Sivertsen 1990), in German (Luukkonen, Persson & Sivertsen 1991a), and in English (Luukkonen, Persson & Sivertsen 1991b). We would also publish together in the obscure journal with an orange cover (Luukkonen, Tijssen, Persson & Sivertsen 1994).

As Olle noted in his presentation at the conference in 1988, there was an increasing focus at that time on publication output and citation impact as indicators for research evaluation and policy. But bibliometric studies were mainly performed within the natural sciences. Olle was ahead of time with his paper about the social sciences. Now, a little more than twenty years later, and after the digital revolution, bibliometrics is expected to cover all areas of research.

Some of the pressure in this direction comes from the new models for performance based funding of research institutions. In these models, the performance of institutions as such is measured, and the experience so far (e.g. in Flanders, United Kingdom and Sweden) shows that it is politically impossible to leave out the social sciences and the humanities. Two or three alternative – or combined – solutions are therefore discussed (Hicks & Wang 2009): One of them is to rely on "the recent aggressive expansion by WoS and Scopus", which points in the direction of full coverage of "all sound journals". The second "is hinted at in two current metrics-based systems, the Norwegian and Australian. Both rely on national research documentation systems." The third is "creating an electronic, full text infrastructure for European SSH literature".

For the discussion of the alternatives, more information seems to be needed about the current publication practices in different fields. The Norwegian system (Sivertsen 2008) so far provides complete data for 30.000 scientific publications (fractionalized counts) from Norway's higher education sector in four years, 2005-2008. I have analyzed the data in a simple manner in *table 1*. It shows field variation in publication practices in three dimensions: Coverage by *Web of Science* (WoS), use of foreign language (versus Norwegian), and publication type (articles in journals and series, articles in books or proceedings, and books). Articles in series with an ISSN are counted as journal articles. All publication counts have been fractionalized between the authors and their institutions. The counts are limited to publications that have appeared with a scientific or scholarly content and format in a publication channel with peer review, but publications in local publication channels (with more than two thirds of the publishing authors are from the same institution) are not counted.

The variations in publication patterns are shown as percentages of the total of publications within each subfield and major field. The analysis shows that *WoS* currently covers two thirds of the scientific journal articles from the higher education institutions in Norway. If books and articles in books are also considered, *WoS* covers about half of the total output. But these shares show large variations, not only between the major fields, but also within them. The large variations within the humanities and social sciences indicate that publication patterns differ with the aims and the subject matter of research, and that it is difficult to point at a certain publishing practice as a quality standard. Although new results from research generally need to be exposed to criticism and further use among the widest possible audience of experts, it is not necessarily a sign of higher quality that the publication is a journal article in an international language, and that this article is indexed for a certain database.

A journal for sociologists publishing in Swedish, *Sociologisk Forskning*, has for a long time been covered by the WoS. The parallel journal for political scientists in Sweden, *Statsvetenskaplig Tidsskrift*, is not covered by WoS. This may not seem to make much difference, since there are so many other journals in political science that are covered by WoS, and since it is generally agreed that the social sciences should strive for internationalization. But from our Norwegian data, we know that while international publishing – from one country's point of view – is widely dispersed among many publication channels, publishing on the national level is concentrated in a few channels that have very few publications from other countries. In Norwegian sociology, 54 per cent of the journal articles are concentrated in only four national journals. The other 46 per cent are dispersed among 57 other journals (including *Acta Sociologica*). In political science, 27 per cent of the articles are concentrated in only two national journals. The rest of the articles are published in 121 other journals. Such skewed distributions are even more apparent when we look at book publishing.

Since publication patterns vary more across disciplines than across countries, I expect that we will see similar skewed distributions in other countries as well. This means that the omission or addition in a central international database of one national publication channel, e.g. *Sociologisk Forskning* or *Statsvetenskapelig Tidsskrift*, will have great effect on the measurement of the publication output from a national point of view. But from an international point of view, the effect on the overall coverage and indicators will be marginal or almost invisible. Bradford's law and the notion of "core journals" can continue to be the cornerstones of coverage policy. It can even be argued, as I once did, that the addition of internationally insignificant journals may distort international comparisons (Sivertsen 1992). But for other purposes, publication channels that are significant on national level do indeed represent a coverage problem.

As noted above, bibliometric databases are now expected to fulfill such other purposes. In response, *Scopus* (Elsevier) and *Web of Science* (Thomson Reuters) are competing and expanding by covering more journals, more proceedings, and even books. The picture I show in two of the columns in *table 1* is becoming historic. WoS coverage is increasing and Scopus is expanding maybe even beyond WoS. It will be interesting to see how WoS and Scopus will meet the challenges that can be seen in the four other columns of *table1*. It will also be interesting to see if there will be alternative or supplementary ways of meeting these challenges. Anyhow, the coverage of the major international bibliographic databases is probably no longer only the question of how to combine Bradford's law with market opportunities and profit margins. Table 1. Braun Score values (in per cent) for some scientometricians

References

Hicks, Diana and Jian Wang. 2009. Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project. Unpublished report, School of Public Policy, Georgia Institute of Technology, April 2009.

Luukkonen, Terttu, Olle Persson and Gunnar Sivertsen. Et "verdenskart" over internasjonalt forskningssamarbeid. Forskningspolitikk (Oslo), no. 3-4/1990, p. 12-14.

Luukkonen, Terttu, Olle Persson and Gunnar Sivertsen. 1991a. Internationale wissenschaftliche Kooperationsnetze. In: P. Weingart et al. (Hg.): Indikatoren der Wissenschaft und Teknik. Theorie, Methoden, Anwendungen, Campus Verlag, Frankfurt am Main, S. 11-33.

Luukkonen, Terttu, Olle Persson and Gunnar Sivertsen.1991b. Nordic Collaboration in Science - a bibliometric study.Nord 1991:28. Nordic Council of Ministers, Copenhagen.

Luukkonen, Terttu, Robert J. W. Tijssen, Olle Persson and Gunnar Sivertsen. The Measurement of International Scientific Collaboration. Scientometrics 1993, 28, 1, p. 15-36.

Persson, Olle. 1988. Nordisk forskning på den internationella tidskriftsmarknaden innom samhällsvetenskap. In: Elizabeth Lundberg (ed.): Internationell vetenskaplig publicering i Norden, Nordisk Ministerråd, Copenhagen, p. 40-57.

Sivertsen, Gunnar. 1988. Internasjonal markedsføring av nordiske forskningspublikasjoner. In: Elizabeth Lundberg (ed.): Internationell vetenskaplig publicering i Norden, Nordisk Ministerråd, Copenhagen, p. 58-67.

Sivertsen, Gunnar. 1992. Should a new bibliometric database for international comparisons be more restricted in journal coverage? In: A.J.F. van Raan et al. (ed.): Science and Technology in a Policy Context. Select Proceedings of the joint EC - Leiden Conference on Science & Technology Indicators, Leiden 1991, DSWO Press, Leiden 1992, p. 35-50.

Sivertsen, Gunnar. 2008. Experiences with a bibliometric model for performance based funding of research institutions. In: J. Gorraiz J. and E. Schiebel (eds): Excellence and Emergence. A New Challenge for the Combination of Quantitative and Qualitative Approaches. Book of Abstracts. 10th International Conference on Science & Technology Indicators. Vienna, Austria. 17-20 September 2008. Vienna (AUT): Austrian Research Centers GmbH, pp. 126-131.

Address of congratulating author:

GUNNAR SIVERTSEN
NIFU STEP, Wergelandsveien 7, N-0167 Oslo (Norway).
E-mail: gunnar.sivertsen@nifustep.no

Table 1 Distribution of scientific publications in Norway's Higher Education Sector 2005-2008 according to coverage by Web of Science, use of foreign language and publication type.

Major field	**Subfield**	**WoS coverage of all publications**	**WoS coverage of journal articles**	**Foreign language**	**Articles (ISSN)**	**Articles in books (only ISBN)**	**Books**
Engineering	Engineering	63 %	86 %	97 %	74 %	26 %	0 %
Health	Biomedicine	97 %	98 %	100 %	98 %	2 %	0 %
	Clinical Medicine	94 %	95 %	83 %	99 %	1 %	0 %
	Dentistry	57 %	57 %	64 %	99 %	1 %	0 %
	Neurology	95 %	99 %	99 %	97 %	3 %	0 %
	Nursing Sciences	40 %	47 %	54 %	86 %	14 %	1 %
	Pharmacology and Toxicology	88 %	91 %	93 %	98 %	2 %	0 %
	Psychiatry	79 %	84 %	92 %	94 %	5 %	1 %
	Psychology	49 %	65 %	68 %	76 %	22 %	2 %
	Social Medicine	63 %	72 %	79 %	87 %	12 %	1 %
	Social Work and Health Care	9 %	20 %	33 %	43 %	51 %	6 %
	Sports Sciences	62 %	79 %	91 %	79 %	20 %	0 %
	Surgery	93 %	96 %	100 %	97 %	3 %	0 %
	Veterinary Sciences	87 %	88 %	89 %	98 %	2 %	0 %
Health	All subfields	75 %	84 %	81 %	90 %	9 %	1 %
Humanities	Archaeology	11 %	22 %	50 %	49 %	47 %	4 %
	Architecture and Design	5 %	8 %	44 %	59 %	36 %	6 %
	Art History	9 %	18 %	44 %	51 %	39 %	10 %
	Asian and African Studies	9 %	21 %	89 %	45 %	48 %	7 %
	Classical Studies	7 %	11 %	50 %	65 %	31 %	4 %
	English Studies	18 %	51 %	86 %	35 %	59 %	6 %
	Ethnology	4 %	9 %	34 %	47 %	46 %	7 %
	Gender Studies	6 %	14 %	31 %	43 %	56 %	1 %
	Germanic Studies	10 %	27 %	96 %	38 %	54 %	8 %
	History	16 %	33 %	36 %	48 %	46 %	7 %
	Linguistics	21 %	36 %	75 %	59 %	38 %	3 %
	Literature	10 %	17 %	28 %	58 %	39 %	3 %
	Media and Communication	3 %	8 %	52 %	38 %	55 %	8 %
	Music	8 %	16 %	34 %	51 %	45 %	5 %
	Philosophy	7 %	12 %	38 %	58 %	34 %	9 %
	Religion and Theology	7 %	14 %	39 %	48 %	45 %	7 %
	Romance Studies	18 %	45 %	82 %	40 %	51 %	9 %
	Scandinavian Studies	0 %	1 %	12 %	30 %	64 %	6 %
	Slavic Studies	6 %	12 %	86 %	50 %	44 %	7 %
	Theatre Studies	9 %	14 %	50 %	60 %	39 %	2 %
Humanities	All subfields	9 %	18 %	44 %	47 %	47 %	6 %

Major field	Subfield	WoS coverage of all publications	WoS coverage of journal articles	Foreign language	Articles (ISSN)	Articles in books (only ISBN)	Books
Natural	Biology	85 %	89 %	97 %	96 %	4 %	0 %
	Chemistry	96 %	99 %	100 %	97 %	3 %	0 %
	Geosciences	92 %	96 %	99 %	95 %	4 %	0 %
	Informatics	22 %	55 %	93 %	40 %	59 %	1 %
	Mathematics	75 %	85 %	96 %	88 %	11 %	1 %
	Physics	94 %	96 %	99 %	97 %	3 %	0 %
Natural	All subfields	81 %	90 %	97 %	90 %	10 %	0 %
Social	Anthropology	12 %	22 %	65 %	56 %	37 %	7 %
	Business and Administration	18 %	32 %	61 %	58 %	38 %	4 %
	Economics	55 %	69 %	78 %	80 %	18 %	1 %
	Educational Research	7 %	14 %	33 %	49 %	45 %	5 %
	Geography	35 %	44 %	76 %	78 %	19 %	2 %
	Law	2 %	3 %	27 %	64 %	29 %	8 %
	Library and Information Science	33 %	39 %	93 %	85 %	14 %	1 %
	Political Science	27 %	60 %	64 %	45 %	51 %	4 %
	Sociology	12 %	26 %	39 %	45 %	50 %	6 %
Social	All subfields	18 %	30 %	49 %	60 %	36 %	5 %
All fields	All subfields	48 %	67 %	71 %	72 %	25 %	3 %

A Webometric Analysis of Olle Persson

Mike Thelwall

Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

Abstract

This chapter is a webometric analysis of leading scientometrician Olle Persson: a systematic compilation of evidence of his research impact drawing only upon web sources. Whilst a purely bibliometric analysis of his refereed journal articles would demonstrate his mainstream intellectual impact, web-based measures can potentially reveal a wider impact, including within education. Although limitations of the data collection method restricted the analysis to just the impact of Olle's department Inforsk and his bibliometric software Bibexcel, the results showed a wide international impact for both. According to the web data, Bibexcel is the product for which Olle's department is best known and is widely used and recommended in education and research. The production and support of Bibexcel is an unusual research activity that would not be fully appreciated in a bibliometric analysis but is a major achievement that marks Olle as a highly influential information scientist.

Introduction

Olle Persson, a widely published scientometrician, has contributed much more to science than just his articles. As head of the Department of Sociology, Umea University, as author of the bibliometric software Bibexcel (Persson, 2009), and as the creator of many maps of disciplines and fields (e.g., Bibliometric maps of research fields, http://www8.umu.se/inforsk/) using Bibexcel and other software (e.g., Persson, 1994), his impact spreads far wider than just the scientific literature. As a consequence, he forms an interesting case study to explore the extent to which the web can provide impact evidence supplementing that available from traditional bibliometric sources, such as the Thomson-Reuters Web of Science.

There have been many previous webometric impact studies, but none focusing on an individual. As far back as 1998, an investigation was published to find out how often five highly-cited information scientists in the U.S. were mentioned on the web, and why. This study found mentions (or "invocations") using name queries in commercial search engines. The results revealed a wide range of different reasons for mentioning the academics' names online, including conference information and resource guides (Cronin, Snyder, Rosenbaum, Martinson, & Callahan, 1998). Although there have apparently been no person-centred webometric studies since 1998, other than co-authorship maps (Kretschmer, 2004; Kretschmer & Aguillo, 2004), web impact methods have been developed since then and applied to web sites, web pages and ideas (Ingwersen, 1998; Thelwall, Vann, & Fairclough, 2006). These are partially encapsulated in a "web impact report" or a "link impact report" which contain a variety of statistics derived from mentions of one or more phrases or identifying links to one or more web sites, respectively. These have been prepared for organisations including the BBC World Service Trust, the UK's National Endowment for Science, Technology and Innovation Research and the United Nations Millennium Development Programme. The reports are also used to evaluate online resources like digital libraries and digital archives (Zuccala, Thelwall, Oppenheim, & Dhiensa, 2007).

This chapter draws upon web impact report and link impact report methods in an attempt to identify as much evidence as possible about the online impact of Olle Persson as a way of quantitatively revealing aspect of his impact. Whilst link and web impact reports typically are focused on a predefined set of web sites or search phrases, this ego-centred report has a wider focus: anything online connected with Olle Persson. In contrast, the report does not use the set of comparators employed in the former types of report because the objective is not to evaluate Olle's impact but to describe it.

Data and Methods

Scientometric analyses are often constrained by the ease with which data can be accessed. An analysis is often possible if there is a pre-existing database such as the Science Citation Index used by the Web of Science. For a webometric analysis the web itself is the de-facto database but analyses of the whole web are normally constrained to the kind of information that can be extracted from queries submitted to commercial search engines. In this study, two kinds of queries are considered: link searches and text searches.

Link searches are available in the sibling search engines Yahoo! and AltaVista and also, in a truncated form, in Google. Live Search allows a kind of link search not useful here (outlinks). The linkdomain: command allows a search for all links to any page with a given domain name. For example, the command `linkdomain:X` matches all pages containing a link to any page in Olle's Department of Sociology. Adding `-site:X` excludes all pages within Olle's university and so `linkdomain:X -site:X` returns external inlinks or site inlinks. Like citations, these results can be used as an indicator of how much impact the department has and where its impact occurs. It is also possible to search for all pages that link to a given single URL via the link command. For example, `link:X -site:X` in Yahoo! returns a list of pages linking to X. This type of search is useful for a more fine-grained web impact analysis.

Text searches are simple search engine text-based queries. Whilst they are, in principle, a good way to find out how and where an individual is mentioned online, in practice they only work well for unambiguous queries. For Olle Persson, the problem is that both Olle and Persson are common Swedish names and so many or most results for an `"Olle Persson"` search would be about different Olles. Similarly, a search for `"Persson, O"` returns many irrelevant matches. Although in principle it would be possible to manually filter out the irrelevant matches, this is impractical due to the large numbers involved. An alternative strategy is to identify words closely associated by Olle and with search results dominated by him. Two such words are Inforsk, the research group of which Olle is a founder member, and Bibexcel, Olle's free scientometric analysis software. Whilst an analysis of these gives only a partial picture of Olle's work, it is likely to be a useful complement to a standard scientometric analysis and so cast light on an aspect of Olle's work that would not normally be investigated with metrics.

The methods used and reported below are Web Impact Analysis and Link Impact Analysis (Thelwall, 2009). The essential points of these are described below alongside the results.

Results

Digital library mentions of Bibexcel and Inforsk

The traditional way of assessing impact is to count citations using the Science Citation Index or a similar source. The Web of Knowledge (WoK) was used instead for Inforsk to assess its impact (2 April 2009). A search for Inforsk in the address field yielded 27 articles having a very impressive total of 368 citations (Citation Report Address=(inforsk) Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S) and peaking in 2007. The top-cited article had 100 citations: *Understanding Patterns of International Scientific Collaboration* (1992) by Luukkonen T, Persson O, Sivertsen G, Science Technology & Human Values Volume: 17 Issue: 1 Pages: 101-126. This information suggests a narrow, deep impact for Inforsk: producing a moderate number of articles, and these articles being excellent.

This information was complemented with a Google Scholar keyword search for Inforsk, returning a total of 263 research mentions. (2 April 2009). This suggests a much wider impact than the WoK results. Some was due to BibExcel and other results were due to the inclusion of a wider class of documents, such as research reports.

Bibexcel was also explicitly searched for in WoK, but it only yielded 6 results – bibliometric articles written by various authors using Bibexcel for their analysis. These articles were cited only 6 times in total. Google Scholar produced many more mentions, probably because of its full-text search facility. It found 185, of which 9 were citations of Bibexcel in various forms. Eliminating these, Google Scholar reported the existence of at least 176 research documents mentioning Bibexcel. (2 April 2009). For this purpose it seems best to disregard the WoK results and use the Google Scholar results as evidence of significant academic use for Bibexcel.

Quantitative overview of web mentions of Bibexcel and Inforsk

In order to gather evidence of web mentions of Bibexcel and Inforsk, keyword searches for both terms were submitted separately to Google, Yahoo! and Live Search. The professional version of LexiURL Searcher was used with the standard query splitting option switched on. This program submits queries to search engines via their Applications Programming Interfaces, which allow automatic searches. The query splitting option allows extra results to be obtained beyond the normal maximum of 1,000 URLs per search (Thelwall, 2008). LexiURL Searcher was then used to combine the results of all three search engines, eliminating duplicate URLs and counting the number of unique URLs, domains, web sites (identified by domain name ending, e.g., Microsoft.com), Top-Level Domains (TLDs) (e.g., .com, .org, .uk) and second or top-level domains (STLDs) (e.g., .ac.uk, .com.pk, .fr). The statistics generated are summarised in Table 1. Note that the most useful figure is probably the number of domains, because content can sometimes be replicated across different pages of a domain, generating multiple matching URLs from a single cause. Full quantitative results are available at http://cybermetrics.wlv.ac.uk/audit/persson/.

Table 1 shows that Inforsk is more widely mentioned online than Bibexcel but both are mentioned hundreds of times, indicating significant impact.

Table 1 Summary of web matches for Bibexcel and Inforsk.

Base query	URLs	Domains	Sites	STLDs	TLDs
bibexcel	419	247	207	51	39
inforsk	1303	560	465	80	63

Figure 1 summarises the main locations of interest in Bibexcel and Inforsk, as judged by web mentions. Ignoring the generic TLDs, it is clear – and somewhat surprising – that both are extremely international in nature and not just known in Sweden, Scandinavia or even Europe. U.S. universities (.edu), Germany, Spain, Finland, the UK and Denmark also feature prominently. Given the relative size of these countries (and Norway) it still seems that Scandinavia is the region where Bibexcel and Inforsk are best known, however.

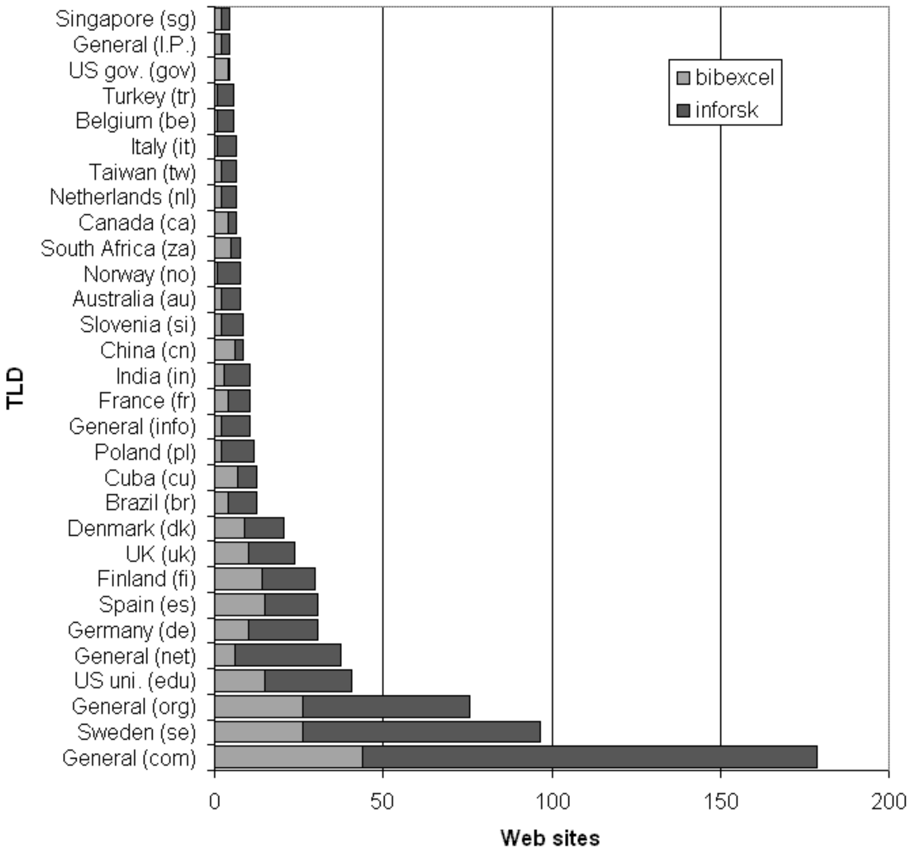

Figure 1 The Top Level Domains containing the most websites mentioning Bibexcel or Inforsk.

Content analysis of invocations of Bibexcel and Inforsk

In order to discover why Inforsk and Bibexcel were mentioned online, an informal content analysis was conducted on 100 random URLs from the combined search results list, using the lists generated by LexiURL Searcher reports. These lists contain a maximum of one URL per domain name to avoid the results being dominated by individual web domains. The content analysis was inductive: starting with a predefined list of categories and adding extra categories when common new types of web mention were found.

Figure 2 reports the results of the content analysis of 100 random web sites mentioning Inforsk. Just over a third of these sites list Inforsk in a list of academic departments. Many of these lists are copies of the Dmoz Open Directory project category: *Science: Science in Society: Academic Departments*. Just under a third of online mentions of Inforsk occur in the context of Bibexcel – mentioning or crediting the department as the host for this software. Just under a fifth of the mentions derive from an Inforsk-authored article in a digital library or hosted on a web site elsewhere on the web. The results overall suggest that Bibexcel is the major product of Inforsk, perhaps even more significant than its research papers. This relative importance is only a tentative suggestion, however, because the content analysis is based upon a maximum of one page per web site and some digital library web sites contain or list many Inforsk-authored articles (e.g., Google Scholar, Ingenta, Wiley Interscience).

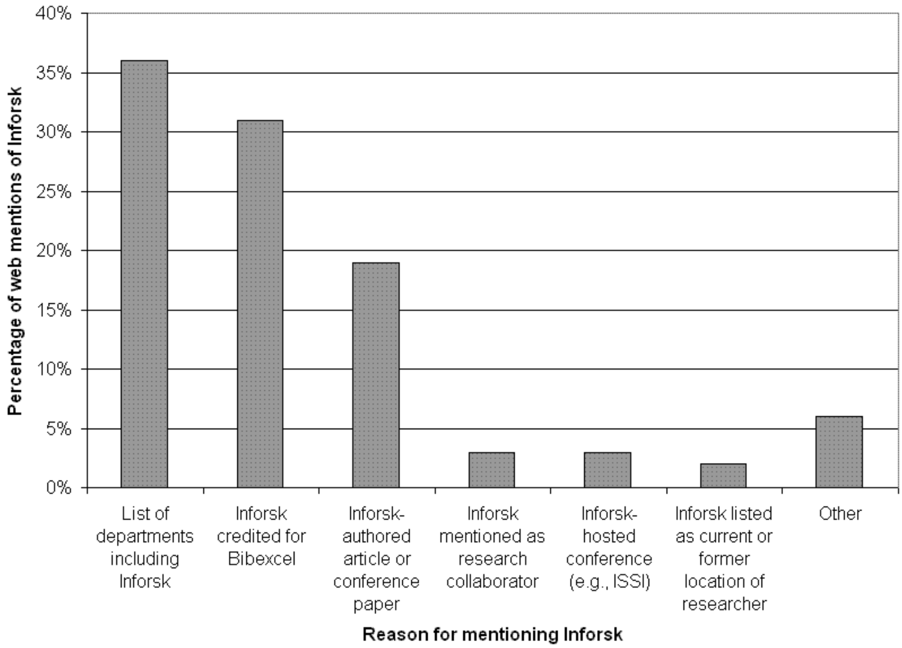

Figure 2 Reasons for mentioning Inforsk in 100 random web sites.

Figure 3 reports the content analysis of 100 random web sites mentioning Bibexcel. The majority of the contexts, almost two thirds, are academic publications: journal articles, conference papers or research reports. Almost all of these papers use Bibexcel to process their data. This shows genuinely extensive use of Bibexcel, especially given that some of the digital library web sites mentioned contain multiple articles mentioning Bibexcel. In addition to this core use of Bibexcel many web site list it as a useful resource, either in a link list or as part of a discussion of how to conduct bibliometric analyses.

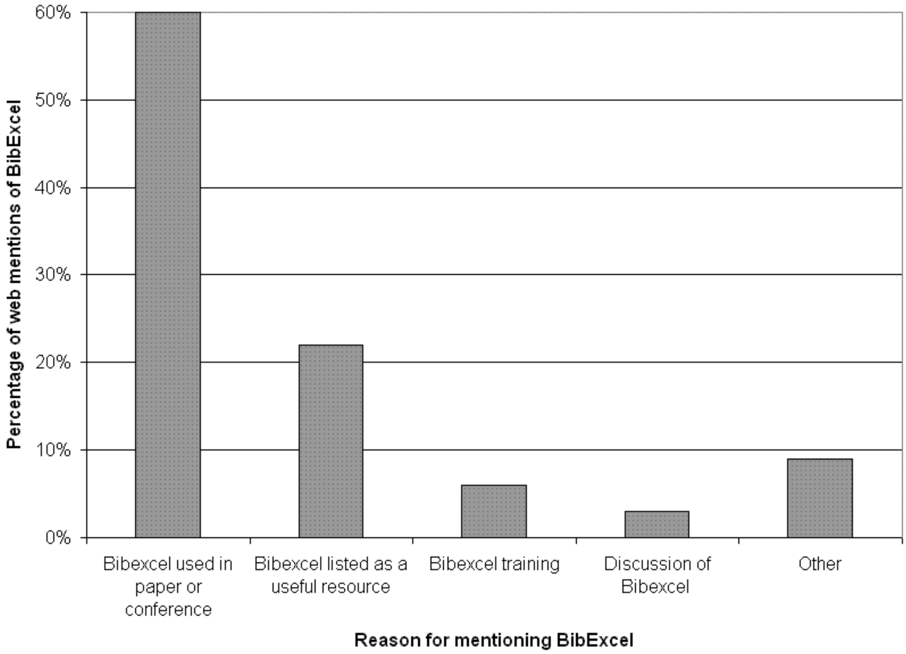

Figure 3 Reasons for mentioning Bibexcel in 100 random web sites.

Conclusion

The web mention data combined from Google, Yahoo! and Live Search clearly shows that Bibexcel and Inforsk are widely known and mentioned online, with hundreds of web sites invoking both. The spread of impact is also international, extending significantly beyond Scandinavia to the rest of Europe and the U.S.

The content analysis of web mentions of Inforsk reveals that it is most frequently mentioned as part of a list of similar departments originating in the open directory dmoz. The second most common cause is Bibexcel, with Inforsk credited as originating the software. This highlights the importance of Bibexcel for Inforsk. The content analysis of Bibexcel reveals it to be commonly credited for use in academic and other research and also often listed as a useful free bibliometric analysis tool. Hence it is clear that Bibexcel is widely seen as a valuable and practical tool for bibliometrics.

In terms of the contribution of Olle Persson, although for technical reasons it was not possible to directly measure the full impact of his ideas, the results clearly show a man with an extraordinary impact on research in terms of his department and software. Bibexcel was an inspired idea that has proven useful to hundreds of researchers around the world and has therefore been an important contribution to scientometrics.

Appendix

Table 2 Top-level domains mentioning Bibexcel or Inforsk.

domain	bibexcel	inforsk	total
com	44	135	179
se	26	71	97
org	26	50	76
edu	15	26	41
net	6	32	38
de	10	21	31
es	15	16	31
fi	14	16	30
uk	10	14	24
dk	9	12	21
br	4	9	13
cu	7	6	13
pl	2	10	12
info	2	9	11
fr	4	7	11
in	3	8	11
cn	6	3	9
si	2	7	9
au	2	6	8
no	1	7	8
za	5	3	8
ca	4	3	7
nl	2	5	7
tw	2	5	7
it	1	6	7
be	1	5	6
tr	1	5	6
gov	4	1	5
I.P.	2	3	5
sg	2	3	5
cz	0	5	5
ar	4	1	5
il	0	4	4
mx	3	1	4
nu	1	3	4
is	1	3	4
ru	1	3	4
us	0	3	3
eu	0	3	3
th	1	2	3
jp	2	1	3
kr	0	3	3
hu	0	2	2
ir	0	2	2
my	0	2	2
co	1	1	2
ch	0	1	1
ro	0	1	1
gr	0	1	1
hr	1	0	1
bg	0	1	1
ua	0	1	1
at	0	1	1
zm	0	1	1
pe	0	1	1
tv	0	1	1
id	0	1	1
yu	0	1	1
ag	0	1	1
np	0	1	1
lv	0	1	1
mt	0	1	1
cg	0	1	1

Reference

Cronin, B., Snyder, H. W., Rosenbaum, H., Martinson, A., & Callahan, E. (1998). Invoked on the web. Journal of the American Society for Information Science, 49(14), 1319-1328.

Ingwersen, P. (1998). The calculation of Web Impact Factors. Journal of Documentation, 54(2), 236-243.

Kretschmer, H. (2004). Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web. Scientometrics, 60(3), 409-420.

Kretschmer, H., & Aguillo, I. F. (2004). Visibility of collaboration on the Web. Scientometrics, 61(3), 405-426.

Persson, O. (1994). The intellectual base and research fronts of JASIS 1986-1990. Journal of the American Society for Information Science, 41(5), 31-38.
Persson, O. (2009). Bibexcel. A tool-box developed by Olle Persson, Inforsk, Umeå univ, Sweden. Retrieved April 15, 2009 from: http://www.umu.se/inforsk/Bibexcel/.
Thelwall, M. (2008). Extracting accurate and complete results from search engines: Case study Windows Live. Journal of the American Society for Information Science and Technology, 59(1), 38-50.
Thelwall, M. (2009). Webometrics.New York: Morgan & Claypool.
Thelwall, M., Vann, K., & Fairclough, R. (2006). Web issue analysis: An Integrated Water Resource Management case study. Journal of the American Society for Information Science & Technology, 57(10), 1303-1314.
Zuccala, A., Thelwall, M., Oppenheim, C., & Dhiensa, R. (2007). Web intelligence analyses of digital libraries: A case study of the National electronic Library for Health (NeLH). Journal of Documentation, 64(3), 558-589.

Address of congratulating author:

MIKE THELWALL
Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton,
Wulfruna Street, Wolverhampton WV1 1LY (UK)
E-mail: m.thelwall@wlv.ac.uk

Pennants for Strindberg and Persson

Howard D. White

College of Information Science and Technology, Drexel University, Philadelphia, PA 19104 (USA)

Abstract

My contribution to the Persson festschrift is another installment, with new data, in my current research program linking (1) relevance theory from linguistic pragmatics with ideas from (2) bibliometrics and (3) information retrieval. Pennant diagrams are a visualization technique I created to bring all three together. I interpret two interesting pennants substantively and conclude with some technical details on creating them.

Why These Two

The title indeed links the renowned Swedish playwright and the honoree of this festschrift, and, yes, it has a personal significance. While I was in Stockholm for the 2005 meeting of the International Society for Scientometrics and Informetrics, I visited the Strindberg Museum (Strindbergs-Museet) and talked about it with Olle at the conference dinner. After I had returned to the States, he sent me a memento—the following quotation from the beginning of Strindberg's novel, *The Red Room,* which is subtitled *Scenes of Artistic and Literary Life*:

> Now the bells of Santa Katrina chimed seven and were echoed by Santa Maria's reedy treble, the Abbey and the German church joined in with their basses, and soon the whole air vibrated with the city's seven bells. And as, one after another, they fell silent, the last one could still be heard in the distance, singing its peaceful evensong. This had a higher note, a purer ring and a swifter tempo than the others....there in the Santa Klara churchyard, whence the bell could still be heard...
>
> TIME: 1879
> PLACE: Stockholm
> CIRCUMSTANCE: A May evening at seven o'clock.
>
> —August Strindberg, *The Red Room*, trans. Elizabeth Sprigge, J. M. Dent & Son, London, 1967, (Everyman Edition), p. 3.

Olle added, "Santa Klara is the church just above Scandic Hotel Continental where we stayed." He also attached a sepia illustration of the Red Room itself as it looked in the 1870s—a hall packed with diners under large chandeliers in Berns Restaurant in central Stockholm. So he and Strindberg remain associated in my memory—perhaps a connection not many others would make!

Introduction

Nevertheless, in bibliometrics even the most dissimilar authors can yield similar patterns of data at the statistical level. As I mulled over ideas for this contribution, it occurred to me that, since both Strindberg and Persson have long records as cited and co-cited authors, both could be mapped in pennant diagrams, a new kind of visualization I introduced in White (2007a, b). Pennants are intended as one demonstration of the explanatory power of Dan Sperber's and Deirdre Wilson's (1995) relevance theory (RT) when it is applied to information science. I thought it highly likely that pennants for Strindberg and Persson as co-cited authors would once again exhibit the relations I had found for other co-cited authors in earlier studies. The abstract of White (2009)—a paper produced just before this one—states that "A central idea in D. Sperber & D. Wilson's relevance theory is that an individual's sense of the relevance of an input in a context varies directly with its cognitive effects and inversely with its ease of processing in that context." The paper goes on to make the nonobvious claim that is explored at length in White (2007a, b): "[A] formula used in information science for weighting search terms in relevance rankings

*Weight = term frequency * inverse document frequency*

instantiates a central idea of Sperber & Wilson's relevance theory from linguistic pragmatics

Relevance = cognitive effects / processing effort.

In other words, cognitive effects and processing effort, which S&W discuss almost exclusively as subjective experiences in individuals, have an objective analogue in the tf*idf formula at the heart of classic information retrieval."

The crisp definition of relevance as an effects/effort ratio is drawn from Goatly (1997). But it is licensed by S&W in many places. For example, Wilson (2007) uses this formulation in a course she gave on relevance theory at the University of London (boldface hers):

- **Revance to an individual**
- Other things being equal, the greater the cognitive effects (of an input to an individual who processes it), the greater the relevance (to that individual at that time).

- Other things being equal, the smaller the processing effort required to derive these effects, the greater the relevance (of the input to that individual at that time).

Pennants are a means of rendering these relations visually. I will defer until the last section a discussion of some technical details that underlie them. My present goal is simply to show how the pennants for my two authors make qualitative sense, starting with Figure 1.

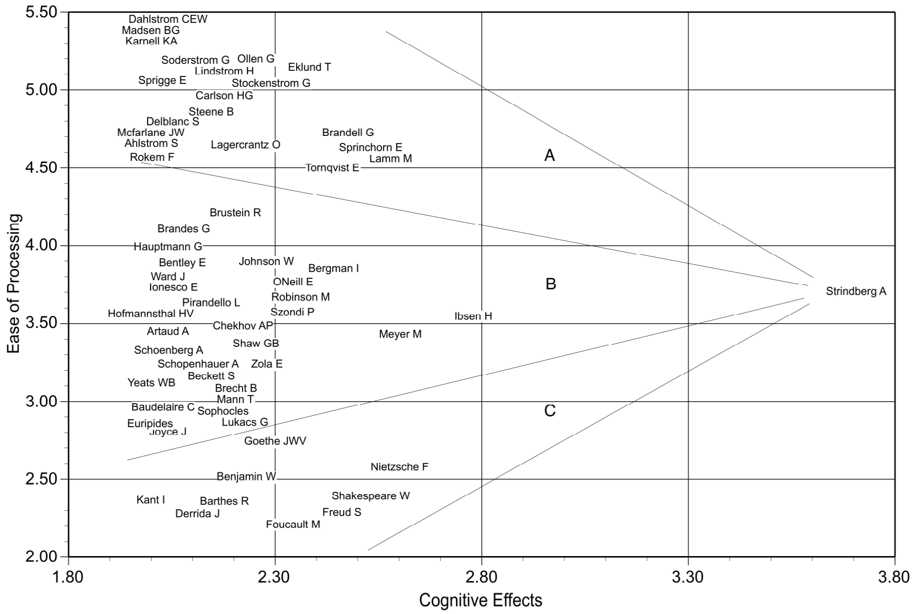

Figure 1 Pennant for authors co-cited at least 10 times with August Strindberg in Arts & Humanities Search on Dialog, April 2009

Strindberg

The pennant is formed by using Strindberg's name as a seed to set an overall context, and the names of authors co-cited with him are, in the language of RT, assumptions in that context. They are not the assumptions of a human mind; rather, they are latent in bibliographic records and made manifest as predictions by an algorithm. Yet all the pennants I have seen exhibit a kind of low-grade artificial intelligence. For example, if one asks literary people to name who immediately comes to mind when "Strindberg" is given as a stimulus, my guess is that the great majority would answer "Ibsen," the other giant of Scandinavian drama. And, sure enough, the rightmost name on the horizontal *cognitive effects* scale is Ibsen's. On the basis of international scholarship, it is Ibsen whose works are predicted to have the greatest cognitive impact when read with

Strindberg's. Note that this *is* only a prediction, not a guarantee. But it accords well with intuition.

More particularly, what "Ibsen H" and "Strindberg A" stand for here is any of their works in any combination. Ibsen's *oeuvre* is large, and Strindberg's is immense (it consists of much more than his plays). However, their *oeuvres* are here being filtered through a mass of co-citing articles that discuss *individual* works. If one drills down to the level of the actual articles, one might find, say, a study comparing *Hedda Gabler* and *Miss Julie*. Therefore, a formidable reading task involving entire *oeuvres* is not necessarily being implied.

Ibsen is the author most highly co-cited with Strindberg, and that is why he has the highest score on the cognitive effects scale. But what does his middling score on the vertical *ease of processing* scale mean? It means that he has been cited in many contexts other than with Strindberg. Pennants always exhibit this structure; the authors most highly co-cited with a seed author are also well cited in other contexts. They tend to have large *oeuvres* with rich implications, both for the seed and beyond. They are thus pulled to the middle of the ease of processing scale. But overall, on grounds of the RT notions of relative cognitive effects and relative processing effort, we can say that Ibsen is most relevant to Strindberg in the pennant diagram.

Discussion of the ease of processing scale brings up the A, B, and C sectors of the pennant. They are in general highly interpretable, both in Strindberg's case and, as will be seen, in Persson's. While the sector lines were drawn by me and not an algorithm, there are good qualitative reasons for putting them about where they are. They reflect differences in the ease of associating authors in the pennant with Strindberg, based on *the specificity of what their works imply* in that context. It is not that any author is more specific than another *as a name*. Rather, as noted, the names of authors designate *oeuvres*, and it is works in those *oeuvres* that differ in the specificity of their relevance to Strindberg studies.

The ease of processing scale is actually based on the idf measure mentioned above. Sparck Jones (1972) created idf as a measure for weighting the "statistical specificity" of terms. The idf measure elevates terms of any sort (here, author names) that occur relatively *infrequently* in a database, because that is taken to indicate specificity. In information retrieval, from which idf comes, more "statistically specific" terms are given higher weights so that documents tagged with them are placed higher in a relevance ranking—the system's prediction that they are more relevant to a query. The opposite is true of terms that occur relatively *frequently*, because such terms are taken to be more general, more nebulous, less indicative of exact content. The idf measure pushes them down in the rankings as probably less relevant to a query.

The idf measure does a similar thing here. Authors cited relatively infrequently, and who are often not well known, are placed high on the ease of processing scale. They turn out to be authors whose works refer to Strindberg or his intellectual world in obvious ways—at the level of titles, subtitles, and chapter headings. In contrast, the more citations authors have, the lower they are placed on the scale and the more famous they are likely to be to domain experts or even the general public. This fame is indicated by the large numbers of citations they have received independently of those they share with the seed author. The idf measure is penalizing the frequent occurrence of their names in the database as if it indicated vagueness and generality. And in a sense it does; their names as cited authors imply countless things. It is this very breadth of implication that makes them relatively hard to relate to the seed.

To get down to cases in Figure 1, the authors in sector A are uniformly associated with critical studies, biographies, and translations of Strindberg. (You can look them up.) While I have not verified their nationalities, I think many are Swedish. To readers not immersed in Strindberg studies, probably all are unfamiliar. I recognized one: Elizabeth Sprigge, who appears as "Sprigge E" in the top left corner. I know her because she is the translator in my Anchor Books edition of Strindberg's plays. The English version of *The Red Room* mentioned at the outset is also hers. Although sector A authors have published books and articles that are easy to relate to Strindberg, they are too specialized to have high citation counts in other contexts, which is why they automatically go to the top here.

Sector B includes many authors who, in contrast to sector A names, are well known indeed. Those who are not may appear in sector B rather than A because their citation counts have been increased by authors whose names are homonyms of theirs. (I did not attempt to disambiguate homonyms.) Michael Robinson, John Ward, and Walter Johnson, for example, are Strindberg scholars whose names in last-name-and-initial style lend themselves to conflation. Michael Meyer is perhaps the best known of all scholars associated with Strindberg and Ibsen; he wrote biographies of both and is a leading translator of Ibsen (by whose name he appears). But "Meyer M," too, could reflect an inflated count. Such doubts aside, it is evident that sector B includes Strindberg's world-class authorial peers. Other than Ibsen, the two pulled nearest to him are, fittingly enough, Ingmar Bergman and Eugene O'Neill. Chekhov, Shaw, and Brecht are close behind; the list of playwrights includes even Sophocles and Euripides. Novelists co-cited with him include Joyce, Mann, and Zola; poets, Yeats and Baudelaire; critics, Robert Brustein, Eric Bentley, and György Lukács. Both Schoenberg, the composer, and Schopenhauer, the philosopher, seem quite comprehensible in this context.

In Sector C some of the most famous authors in the world are mixed with titans of literary fashion. In recent years scholars in the humanities have cited Barthes, Foucault, Derrida, and Walter Benjamin as faithfully as they have drawn breath; as a result, these four are co-cited with practically every artist who ever lived. But looking for felicities, it makes sense that Nietzsche and Freud are predicted to have more cognitive effects than Kant in the context of Strindberg's work, and Shakespeare more than Goethe.

When we get to names of this magnitude, we can draw explicit contrasts with the names in sector A that make the differences in processing them very clear. Which are easier to relate to Strindberg—works by sector A authors like *Strindberg's Ghost Sonata* (Egil Tornqvist), *Strindbergs dramatik* (Gunnar Ollen), and *Strindberg in Inferno* (Gunnar Brandell); or works by Foucault, Goethe, and Nietzsche? Shakespeare is the author of…well, nothing with "Strindberg" in the title. The point is not that the sector C authors are irrelevant to Strindberg; otherwise they would not be co-cited with him. It is that the relevance is much less on the surface; it is literally harder to see.

This shows how the effects/effort ratio from Sperber & Wilson can explain the tf*idf formula used to weight terms in classic document retrieval. The function of idf in relevance rankings of documents is to push *up* documents whose relevance is easy to see and to push *down* documents whose relevance is harder to see. The tf*idf weighting of terms that I used to place authors in the Strindberg pennant is doing much the same thing. It pushes to the top of the pennant names like Tornqvist, Ollen, Brandell, and Sprigge, and to the bottom names like Derrida, Freud, Benjamin, and Kant, even though in both cases the algorithm is completely blind to the qualitative nature of the works each set of names stands for. By the S&W criteria for relevance, works by the authors in sector C are not *ir*relevant but *less* relevant than works by sector A authors, because they require more effort to relate to the seed. Some inquirers will be willing to make this extra effort, but many will be content, if they have questions about Strindberg, to look no further than works like those by the Strindberg experts mentioned above. Retrieval system designers are well aware of this fact, and that is why they use tf*idf and other algorithms like it. Retrievals of obviously relevant documents—*You're interested in Strindberg? Here's some stuff on Strindberg*—make the designers' systems look good to judges in evaluation trials. Nevertheless, one still sees scholars putting considerable effort into the pursuit of less obvious relations, such as the comparative studies implied by the authors co-cited with Strindberg in sectors B and C.

Persson

Figure 2, Persson's pennant, exhibits a similar structure to the one just discussed. It is a bit less symmetrical than Strindberg's because Persson's overall citation count is closer to that of his co-citees in sector A than in sector B. His lower count also results in a shorter horizontal axis. But these minor differences do not affect interpretation.

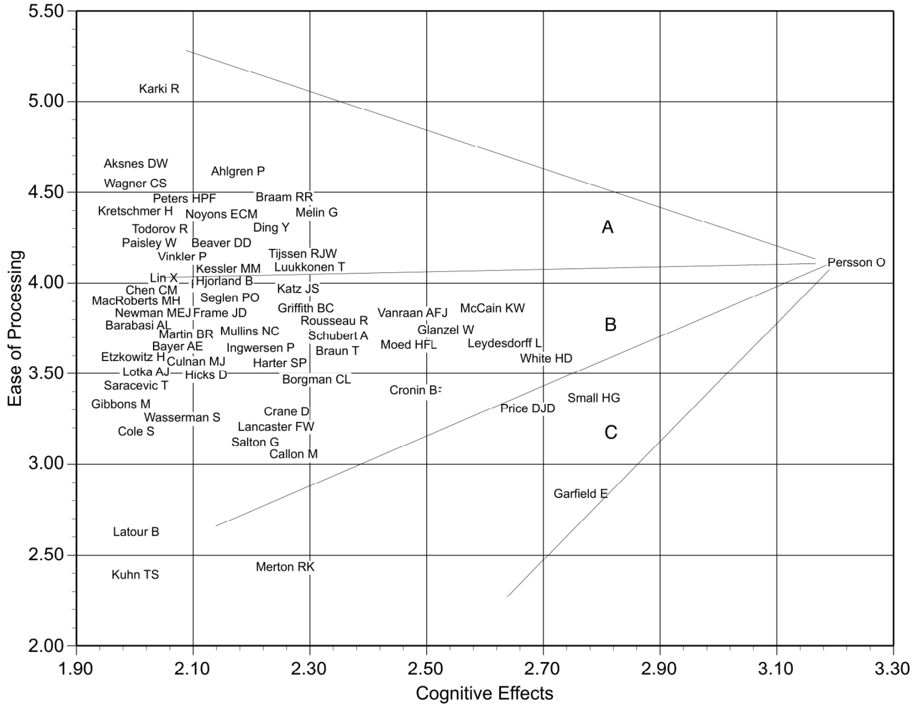

Figure 2 Pennant for authors co-cited at least 10 times with Olle Persson in Social Scisearch on Dialog, March 2009

There may be more European (and Scandinavian) researchers in Persson's pennant than there would be in, say, mine; if so, it seems only natural. Overall, the authors cited with him reflect his identification with bibliometrics as opposed to other specialties in information science. This is not to say he has no ties to information retrieval; one sees, for example, Gerard Salton, F. W. Lancaster, and Tefko Saracevic among his co-citees. But most of the names connote areas of bibliometrics; there is scant evidence of links to research in, say, information behavior or user studies.

The authors nearest Persson on the cognitive effects scale suggest he is most identified with citation analysis. That is what the names of the four authors most co-cited with him—in descending order, Henry Small, Eugene Garfield, myself, and Derek J. de Solla Price—jointly connote. (In the raw data, Small appears as both "Small HG" and "Small H"; I have combined the counts for the two name-forms.) Persson's most cited paper, "The intellectual base and research fronts of JASIS 1986-1990" (1994), is an author co-citation analysis, very much in the line of studies I am continuing here. It is also notable that Katherine W. McCain and Loet Leydesdorff, citation analysts both, are among those with high predicted cognitive effects in the context set by Persson. Most of the other authors in sector B are mainstream information scientists (or crossover figures) whose work reinforces the view of Persson as a bibliometrician (not that anyone doubted it).

Three of Persson's top co-citees in the pennant—Small, Price, and Garfield—are in sector C, implying difficulty in relating their works to his. Again, the idf part of my algorithm registers the fact that Small, Price, and especially Garfield have thousands of citations beyond those they share with Persson. (Garfield's huge body of work is also less topically concentrated than Small's or Price's.) However, if one goes to the level of the articles in which the co-citations occur, this difficulty may be more apparent than real. Compare these three with the other sector C authors at bottom left, Robert K. Merton, Bruno Latour, and Thomas S. Kuhn. Their works differ much more in subject matter from Persson's than the works of Small, Price, and Garfield, and this lessens their predicted cognitive effects. At the same time, these very famous authors are highly cited in numerous disciplines besides information science, and this tells the tf*idf algorithm that they are difficult authors to relate to Persson.

The authors in sector A, in contrast, are easy to relate to him. They differ little from him in research interests. I counted five co-citees who are also his co-authors: Aksnes, Melin, Luukkonen, Tijssen, and Glänzel; all but the last are in sector A. (My line separating sectors A and B is more arbitrary than in Strindberg's case.) Interestingly, Strindberg, too, has some co-authors in sector A of his pennant—not contemporaries of his who share his bylines, but present-day scholars who get title-page credit for editing, translating, or writing introductions to his works. The point of bringing up co-authors is to show, once again, how names placed in sector A imply much the same subjects as the seed author—in the case of co-authors, identical subject matter. Sector A thus represents narrowness of implication, and the other two sectors represent increasing breadth.

A final illustration of high focus in sector A is Riitta Kärki at top left. She has been co-cited with Persson 11 times (just above the threshold of 10 for appearing in the pennant), and so is far from him on the cognitive effects scale. However, she tops the ease of processing scale. Not only is her total citation count quite low, but her co-citations with Persson involve just one article of hers (Kärki 1996) and one article of his (Persson, 1994), both of them author co-citation analyses. This echoes the claim in White (2007a) that "ease of processing" may mean not simply obvious connections of subject matter but also small *oeuvres* and relative brevity of content.

Background Notes

Schneider, Larsen, & Ingwersen (2007), a PowerPoint presentation available on the Web, is a good guide to interpreting pennant diagrams of various kinds. However, the reasoning behind them is complicated, since it must tack between ideas from information science and Sperber & Wilson's relevance theory. Also, some unconventional techniques of analysis are involved. What follows is an attempt to explain these matters in brief.

Pennants are scatterplots of points representing the cognitive effects and processing effort of terms in the context of a seed term. They begin with two sets of frequency counts: (1) the number of times each term in a distribution co-occurs with a seed term, and (2) the number of documents in the database in which each of those terms occurs. The first set of counts is labeled "term frequencies" or tf. The second set is labeled "document frequencies" or df. The tf count is used in a formula to operationalize "cognitive effects" from RT. The df count is used in another formula to operationalize RT's "processing effort."

White (2007a) tells how to obtain both sets of counts from databases on Dialog, which usually is easy to do. It involves forming the set of all documents in which the seed term appears and ranking the terms in those documents with Dialog's Rank command. These are moves in the tradition of Persson's 1986 article on "online bibliometrics," and they can generate all the standard core-and-scatter (i.e., bibliometric) distributions. (It is remarkable that, in many databases, Dialog supplies the exact data needed to produce pennants but makes no further use of them that I am aware of.)

The Dialog results can be copied into DeltaGraph, which is statistical charting software. (Excel, unfortunately, is not usable.) Examples of raw and derived values from a DeltaGraph spreadsheet for the Strindberg pennant appear in Table 1. A judgment sample of authors in sectors A, B and C have been sorted by their values in the "Sector %" column, which conveys the sharp differences in the tf/df ratio over the three sectors. These differences may also often be sensed in the increasing recognizability of author names from Eklund to Foucault.

Table 1 Sample data for making and interpreting the Strindberg pennant

Name	Count with seed	Count overall	Sector %	Log tf	Log idf	Weight
Strindberg A	623	623	100.0	3.79	3.68	13.97
Eklund T	22	25	88.0	2.34	5.08	11.90
Lamm M	38	87	43.7	2.58	4.54	11.71
Brandell G	30	59	50.8	2.48	4.71	11.66
Sprinchorn E	34	80	42.5	2.53	4.57	11.58
Tornqvist E	28	88	31.8	2.45	4.53	11.09

Name	Count with seed	Count overall	Sector %	Log tf	Log idf	Weight
Ibsen H	60	884	6.8	2.78	3.53	9.81
Meyer M	40	1027	3.9	2.60	3.47	9.02
Bergman I	26	578	4.5	2.41	3.72	8.97
Robinson M	23	554	4.2	2.36	3.73	8.82
ONeill E	22	533	4.1	2.34	3.75	8.79
Szondi P	22	836	2.6	2.34	3.55	8.33
Nietzsche F	40	9049	0.4	2.60	2.52	6.56
Shakespeare W	34	12727	0.3	2.53	2.37	6.01
Freud S	29	14547	0.2	2.46	2.31	5.70
Foucault M	22	19213	0.1	2.34	2.19	5.14
	tf	df	(tf/df) * 100	1+Log(tf)	Log(3mil/df)	tf*idf

Manning & Schütze (1999) and Jurafsky & Martin (2000) suggest converting tf and/or idf counts to logarithms to damp the original values. My untested (but plausible) hypothesis is that logarithmic values are truer to our sense of discriminable differences in both the cognitive effects of terms and the effort needed to process them. I use the version of tf*idf weighting given in Manning & Schütze (1999). For the *i*th term in document *j*:

$$\text{weight}_{(i,j)} = (1 + \log(\text{tf}_{i,j}))*\log(N/\text{df}_i)$$

where all term counts ≥ 1, logarithms are base 10, and N is the total number of documents in the collection. In the present study I used 3 million as the value for N in Arts & Humanities Search for Strindberg and Social Scisearch for Persson. The scale values on the axes of both the Strindberg and Persson pennants are base-10 logs.

Multiplying tf by an inverse measure, idf, corresponds to dividing cognitive effects by processing effort. However, since idf values *are* inverse—high when processing effort is low and low when it is high—it reduces mental gymnastics to rename the idf scale "ease of processing"; then high idf means "easy" and low idf means "difficult."

Pennants can be used to show the effect of the tf*idf multiplication—indeed, that was one of the main points of White (2007a)—but it should be noted that they show tf and idf plotted *separately* on the two axes, as in Figures 1 and 2. Pennants thus allow the predicted cognitive effects and processing effort of each data point to be simultaneously read.

It should also be noted that tf is used differently here from its use in information retrieval (IR). Here, it refers to terms in a bibliometric distribution that are rank-ordered by the tf count. There, it is used to weight terms in queries that searchers put to large collections of documents; it designates the number of times each query term occurs in each document. In pennants, the entire set of bibliographic records formed in Dialog is considered one big document, and tf designates how frequently terms in that big document co-occur with the seed term.

The meaning of document frequency or df also differs somewhat in my RT-influenced line of research as against traditional information retrieval. When I use "document frequency" with a bibliometric distribution generated by the Rank command, it refers to how frequently each term in a large distribution of terms occurs in a document in the database. In IR, "document frequency" refers to the number of documents in the collection that contain a given search term.

Even so, I do not see these differences as major, because my purpose in adapting the tf*idf formula to bibliometric distributions is to show that relevance theory can explain its function in information retrieval in a new way. RT holds relevance-seeking to be a basic component of human cognition (Sperber & Wilson, 1995). IR uses the tf*idf formula to rank documents algorithmically by their relevance to a query. If relevance is defined as in the RT ratio in the introduction—as varying directly with the cognitive effects and inversely with the processing effort of a communicative input—then these two variables should be discernible in relevance rankings of documents or terms representing them, and so, in fact, they prove to be.

Since I lacked human relevance judgments of documents to work with, I applied the tf*idf formula (as it appears in Manning & Schütze, 1999) to bibliometric data on documents from Dialog, recalling that the bibliometric distributions have long been considered, as Saracevic (1975) puts it, "relevance-related." What I found, and have repeatedly confirmed, is that the frequencies of term co-occurrences with the seed term (tf) are a promising model of the cognitive effects of those terms in that context. More interestingly, the inverse document frequency (idf) measure is a promising model of the effort of processing the same terms in that context.

Interpretations like mine may seem to read too much into the tf*idf formula, a mechanical procedure. I would counter that the *verbal* parts of bibliometric data (White, 2005) most need detailed, qualitative analysis when complicated and somewhat novel concepts are being presented, as here. Despite my somewhat poetic approach, I think the predictions sketched in my relevance-theoretic work, starting with White (2007a, b), are empirically testable. The testing will probably require someone more grounded in experimental research than I. My goal at present is simply to interest researchers in using bibliometric data psychologically. Relevance theory, which Sperber & Wilson have consciously aligned with cognitive science, seems like a good place to begin looking for theoretical foundations.

References

Goatly, A. (1997). The language of metaphors. London and New York: Routledge.
Kärki, R. (1996). Searching for bridges between disciplines—An author cocitation analysis on the research into scholarly communication. Journal of Information Science, 22, 323-334.
Jurafsky, D., & Martin, J.H. (2000). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River, NJ: Prentice Hall.
Manning, C.D., & Schütze, H. (1999). Foundations of statistical natural language processing. Cambridge, MA: MIT Press.
Persson, O. (1986). Online bibliometrics—A research tool for every man. Scientometrics, 10, 69-75.
Persson, O. (1994). The intellectual base and research fronts of JASIS 1986-1990. Journal of the American Society for Information Science, 45, 31-38
Saracevic, T. (1975). Relevance: A review of and framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, 321-343.
Schneider, J.W., Larsen, B., & Ingwersen, P. (2007). Pennant diagrams, what is it [sic], what are the possibilities and are they useful? Presentation at the 12th Nordic Workshop in Bibliometrics and Research Policy, Copenhagen, Denmark, September 13-14, 2007. Retrieved April 15, 2009 from www.db.dk/nbw2007/files/2c_Peter_Ingwersen.pdf
Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application to retrieval. Journal of Documentation, 28, 11-21.
Sperber, D., & Wilson, D. (1995). Relevance: Communication and cognition. (2d ed.) Oxford, UK, and Cambridge, MA: Blackwell.
White, H.D. (2005). On extending informetrics: An opinion paper. Proceedings of ISSI 2005, the 10th International Conference of the International Society for Scientometrics and Informetrics. Stockholm, Sweden: Karolinska University Press. Vol. 2: 442-449.
White, H.D. (2007a). Combining bibliometrics, information retrieval, and relevance theory. Part 1: First examples of a synthesis. Journal of the American Society for Information Science and Technology, 58, 536-559.
White, H.D. (2007b). Combining bibliometrics, information retrieval, and relevance theory. Part 2: Some implications for information science. Journal of the American Society for Information Science and Technology, 58, 583-605.
White, H.D. (2009). Some new tests of relevance theory in information science. Paper accepted for the Proceedings of ISSI 2009, the 12th International

Conference of the International Society for Scientometrics and Informetrics. (In print.)
Wilson, D. (2007). Relevance: the cognitive principle. Lecture 3 of Pragmatic Theory (PLIN2002) 2007-08. Retrieved April 15, 2009 from http://www.phon.ucl.ac.uk/home/nick/content/pragtheory/PRAG3.doc

---

Address of congratulating author:

HOWARD D. WHITE
College of Information Science and Technology, Drexel University,
Philadelphia, PA 19104 (USA)
E-mail: whitehd@drexel.edu

The Bibliography of Professor Olle Persson

Järvelin, Kalervo; Persson, Olle, The DCI Index: Discounted Cumulated Impact-Based Research Evaluation. // Journal of the American Society for Information Science and Technology. - 2008 (59) : 9, s. 1433-1440

Persson, Olle, Exploring the analytical potential of comparing citing and cited source items. // Scientometrics. - 2006 (68) : 3, s. 561-572

Persson, Olle, Bibliometric survey of Swedish health economics research. // Scandinavian Journal of Public Health : An Evaluation of Swedish Health Economics Research. - 2006 : Supplement 68, s. 21-25

Persson, Olle, Bibliometri - aktuella trender. // InfoTrend : Nordisk Tidskrift for Informationsspecialister. - 2006 (61) : 4, s. 107-111

Cornelius, Barbara; Landstrom, Hans; Persson, Olle, Entrepreneurial Studies: The Dynamic Research Front of a Developing Social Science. // Entrepreneurship Theory and Practice. - 2006 (30) : 3, s. 375-398

Cornelius, Barbara; Persson, Olle, Who's who in venture capital research. // Technovation. - 2006 (26) : 2, s. 142-150

Persson, Olle, "Citation Indexes for Science" - A 50 year citation history. // Current Science. - 2005 (89) : 9, s. 1503-1504

Dastidar, Prabir G; Persson, Olle, Mapping the global structure of Antarctic research vis-à-vis Antarctic Treaty System. // Current Science. - 2005 (89) : 9, s. 1552-1554

Glenisson, Patrick; Glänzel, Wolfgang; Persson, Olle, Combining full-text analysis and bibliometric indicators. A pilot study. // Scientometrics. - 2005 (63) : 1, s. 163-180

Persson, Olle; Glänzel, Wolfgang; Danell, Rickard, Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. // Scientometrics. - 2004 (60) : 3, s. 421-432

Persson, Olle; Danell, Rickard, Decomposing national trends in activity and impact. : A study of Swedish neuroscience papers. // Handbook of Quantitative Science and Technology Research. : The Use of Publication and Patent Statistics in Studies of S&T Systems, Moed, Henk F,; Glänzel, Wolfgang; Schmoch, Ulrich (ed.). - 2004. - S. 515-528

Meyer, Martin; Pereira, Tiago Santos; Persson, Olle; Granstrand, Ove, The scientometric world of Keith Pavitt. A tribute to his contributions to research policy and patent analysis. // Research Policy. - 2004 (33) : 9, s. 1405-1417

Glänzel, Wolfgang; Danell, Rickard; Persson, Olle, The decline of Swedish neuroscience: Decomposing a bibliometric national science indicator. // Scientometrics. - 2003 (57) : 2, s. 197-213

Danell, Rickard; Persson, Olle, Regional R&D activities and interactions in the Swedish Triple Helix. // Scientometrics. - 2003 (58) : 2, s. 205-218

Persson, Olle: All author citations versus first author citations// SCIENTOMETRICS – 2001 (50) 2, s. 339-344

Melin, Goran, Olle Persson & Rickard Danell: A bibliometric mapping of the scientific landscape on Taiwan// ISSUES & STUDIES – 2000 (36) 5, s. 61-82

Persson, Olle: A tribute to Eugene Garfield - Discovering the intellectual base of his discipline// CURRENT SCIENCE – 2000 (79); 5, s. 590-591

Mahlck, Paula & Olle Persson: Socio-bibliometric mapping of intra-departmental networks// SCIENTOMETRICS – 2000 (49); 1, s. 81-91

Persson, O: Luukkonen, T; Hälikkä, S: A Bibliometric Study of Finnish Science – 2000, VTT Group for Technology Studies: Helsinki

Meyer, Martin & Olle Persson: Nanotechnology - Interdisciplinarity, patterns of collaboration and differences in application//SCIENTOMETRICS – 1998 (42); 2, s. 195-205

Beckmann, Martin & Olle Persson: The thirteen most cited journals in economics// SCIENTOMETRICS – 1998 (42); 2, s. 267-271

Danell, Rickard, Lars Engwall & Olle Persson: The first mover and the challenger: The relationship between two journals in organization research// SCIENTOMETRICS – 1997 (40); 3, s. 445-453

Melin, Goran; Persson, Olle: Hotel cosmopolitan: A bibliometric study of collaboration at some European universities// JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE - 1998

Persson Olle; Melin, Goran; Danell, Rickard; Kaloudis, Aris: Research collaboration at Nordic universities// SCIENTOMETRICS – 1997 (39); 2, s. 209-223

Persson, Olle; Melin, Goran: Equalization, growth and integration of science//SCIENTOMETRICS – 1996 (37); 1, s. 153-157

Melin, Goran; Persson, Olle: Studying research collaboration using co-authorships// SCIENTOMETRICS – 1996 (36); 3, s. 363-377

Persson, Olle: ISI miscount?//NATURE - 1996, (380); 6570, s. 100-100

Persson, Olle; Beckmann, Martin: Locating the network of interacting authors in scientific specialties// SCIENTOMETRICS – 1995 (33); 3, s. 351-366

Persson, Olle: The intellectual base and research fronts of JASIS 1986-1990// JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE – 1994 (45); 1, s. 31-38

Luukkonen, T; Tijssen, RJW; Persson, O; Sivertsen, G: The measurement of international scientific collaboration// SCIENTOMETRICS – 1993 (28); 1, s. 15-36

Andersson, Ake E. & Olle Persson: Networking scientists// The Annals of Regional Science – 1993 (27) 1, s. 11-21
Luukkonen, T; Persson, O; Sivertsen, G: Understanding patterns of international scientific collaboration// SCIENCE TECHNOLOGY & HUMAN VALUES – 1992 (17); 1, s. 101-126
Luukkonen, Tertuu, Olle Persson, Gunnar Sivertsen: Nordic Collaboration in Science: A Bibliometric Study – 1991, Nordic Council of Ministers
Broady, Donald; Persson, Olle: Bourdieu in the USA - bibliometric notes// SOCIOLOGISK FORSKNING – 1989 (26); 4, s. 54-73
Hoglund, Lars; Persson, Olle: Communication within a national r-and-d-system - a study of iron and steel in Sweden// RESEARCH POLICY – 1987 (16); 1, s. 29-37
Hoglund, Lars; Persson, Olle: The use of knowledge: a long term research program – 1987, Inforsk: Umeå
Persson, Olle: Online bibliometrics - a research tool for every man// SCIENTOMETRICS – 1986 (10); 1-2, s. 69-75
Persson, Olle: Graphing online searches with Lotus 1-2-3// DATABASE – 1986 (9); 2, s. 57-59
Persson, Olle: Scandinavian social-science in international journals// SOCIAL SCIENCE INFORMATION STUDIES – 1985 (5); 4, s. 185-190
Persson, Olle: Sociologists and their classics// SOCIOLOGISK FORSKNING – 1985 (22); 1, s. 89-90
Ellis, D; Roberts, N; Hounsell, D; Saracevic, T; Persson, O: Information man or information action as a heuristic for information studies - comments on the 2 positions// SOCIAL SCIENCE INFORMATION STUDIES – 1985 (5); 1, s. 25-32
Persson, Olle: Informell kommunikation bland forskare och tekniker – 1980, Sociologiska inst: Umeå universitet
Persson, Olle: Informell kommunikation bland forskare och tekniker – 1980, Sociologiska inst: Umeå universitet
Hoglund, Lars; Persson, Olle: Kommunikation inom vetenskap och teknik – 1980, Reports from the Department of Sociology
Hoglund, Lars; Persson, Olle: Datorbaserad litteratursokning: två fallstudier – 1980, Sociologiska inst: Umeå universitet
Hoglund, Lars; Persson, Olle: Gatekeeperfunktioner och IoD-verksamhet – 1978, Sociologiska inst: Umeå universitet
Hoglund, Lars; Persson, Olle: A survey of studies on use and production of scientific and technical information – 1978, Research reports from the Department of sociology: University of Umeå

Hoglund, Lars; Persson, Olle: Information use within an applied technical field – 1978, Research reports from the Department of sociology: University of Umeå
Hoglund, Lars; Persson, Olle: Informationsutnyttjande inom ett tekniskt specialområde – 1977, Sociologiska inst: Umeå universitet
Hoglund, Lars; Persson, Olle: Evaluation of a Computer-Based Current Awareness Service for Swedish Social Scientists – 1975, Research reports from the Department of sociology: University of Umeå
Hoglund, Lars; Persson, Olle: Evaluation of a Computer-based Current Awareness Service for Swedish Social Scientists – 1975, Dept. of Sociology: University of Umeå
Hoglund, Lars; Persson, Olle: Transportkonsumtionen i sociologisk belysning: en explorativ analys av bilinnehav och bilanvändning – 1973, Sociologiska inst: Umeå universitet

Addendum

Finally, it is our pleasant duty to announce the names of those colleagues and friends of Olle who were not able to contribute to this volume, but who have expressed their wish to congratulate Olle Persson on the occasion of his birthday. Herewith, we kindly acknowledge best wishes expressed by

DAG AKSNES
NIFU-STEP (Norway)

LENNART BJÖRNEBORN
Royal School of Library and Information Science, Copenhagen (Denmark)

BLAISE CRONIN
University of Indiana (U.S.A)

PETER INGWERSEN
Royal School of Library and Information Science, Copenhagen (Denmark)

GÖRAN MELIN
Ministry of Education and Research (Sweden)

ED NOYONS
University of Leiden (Netherlands)

BALÁZS SCHLEMMER
ISSI administration (Hungary)

HENRY SMALL
Thomson Reuters (U.S.A)

ROBERT TIJSSEN
University of Leiden (Netherlands)

IRENE WORMELL
Informatiker Konsult AB (Sweden)

research
bibliometric
Swedish
forskning
Publication
Umeå
Danell
studies
evaluation
Olle
professor
Cornelius
Glänzel
Citation
indicators
Persson
Science
Scientometrics
Information
Library
Sociologiska
School
institutionen
analysis
universitet
health
national