Citation Tracking of Scientific Publications through Two Different Searching Tools: Google Scholar and Web of Science

Maurella DELLA SETA, Rosaria CAMMARANO
Documentation Service
Istituto Superiore di Sanità, Rome, Italy
maurella.dellaseta@iss.it

INSTITUTIONAL ARCHIVES FOR RESEARCH: EXPERIENCES AND PROJECTS IN OPEN ACCESS
Istituto Superiore di Sanità
Rome, 30 November - 1 December 2006

Purpose of this paper

- To compare Google Scholar with Thomson Scientific Web of Science
- Similarities and differences in citation analysis of scientific publications
- Evaluate the performance of these two tools in tracking citations of two predefined sets of publications.

Citation analysis

- Has been used since the mid-20th century as a tool to measure impact and visibility of scientific articles, to monitor a subject trend, and to evaluate the scientific impact of a given researcher or institution
- Eugene Garfield from the Institute of Scientific Information of Philadelphia (ISI) in 1955 developed the idea of constructing a citation index for science, based on the assumption that the more an article is cited in subsequently published papers, the higher its impact on the scientific community

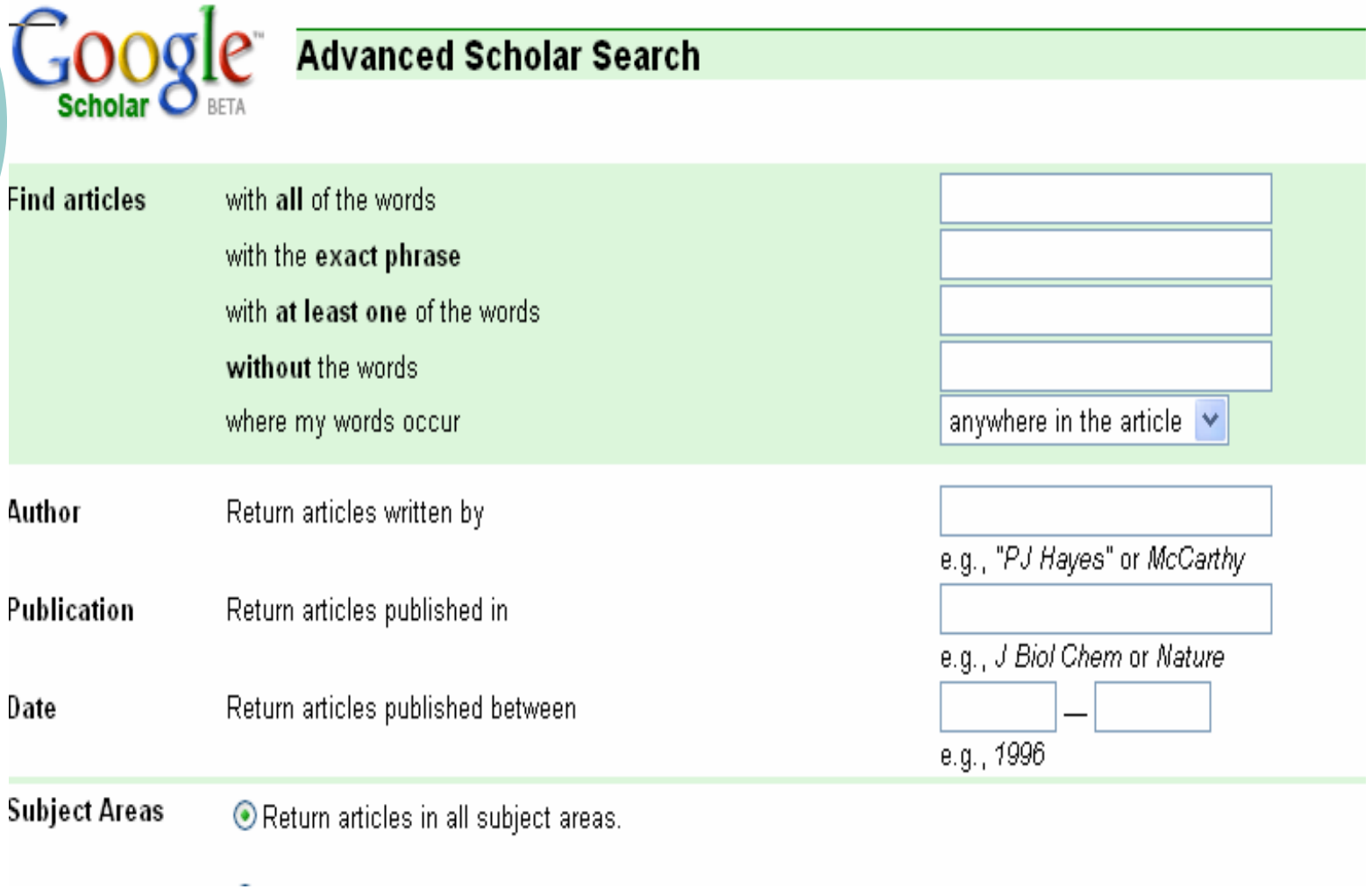
Citation analysis (2)

- Although it has been recognized as an often inappropriately applied practice, the impact factors of journals in which authors publish are still employed in many countries to evaluate the performance of an author or of a research group in order to award grants or government funding

Google Scholar
http://scholar.google.com

- Free-of-charge search engine aimed at finding scholarly information on the Web
- Available in its beta version since November 2004
- Its search includes bibliographic references and the full-text of peer-reviewed papers, theses, books, abstracts, technical reports. It includes PowerPoint presentations and preprints from universities, academic institutions and professional societies.

Google Scholar - Advanced Scholar Search



“Cited by” Search Option

Potential involvement of Fas and its ligand in the pathogenesis of Hashimoto's thyroiditis - group of 4 »
... Richiusa, G Papoff, G Ruberti, M Bagnasco, **R Testi** ... - Science, 1997 - jautoimdis.com
The mechanisms responsible for thyrocyte destruction in Hashimoto's thyroiditis
(HT) are poorly understood. Thyrocytes from HT glands, but not from **...**
Cited by 357 - Related Articles - Cached - Web Search - ACNP Holdings - BL Direct

Requirement for GD3 ganglioside in CD95-and ceramide-induced apoptosis. - group of 2 »
... Agostino, B Tomassini, A Zeuner, MR Rippo, **R Testi** - Science, 1997 - ncbi.nlm.nih.gov
Gangliosides participate in development and tissue differentiation.
Cross-linking of the apoptosis-inducing CD95 protein (also called Fas or **...**
Cited by 206 - Related Articles - Cached - Web Search - ACNP Holdings - BL Direct

The CD69 receptor: a multipurpose cell-surface trigger for hematopoietic cells. - group of 2 »
R Testi, DD'Ambrosio, R De Maria, A Santoni - Immunol Today, 1994 - ncbi.nlm.nih.gov
CD69 was initially described as being restricted to recently activated lymphoid
cells, but is now known to be expressed on the surface of all **...**
Cited by 205 - Related Articles - Cached - Web Search - BL Direct

Multiple pathways originate at the Fas/APO-1 (CD95) receptor: sequential involvement of ... - group of 3 »
... De Maria, G Camarda, A Santoni, G Ruberti, **R Testi** - EMBO J, 1995 - pubmedcentral.nih.gov
The early signals generated following cross-linking of Fas/APO-1, a
transmembrane receptor whose engagement by ligand results in apoptosis **...**
Cited by 171 - Related Articles - Cached - Web Search - ACNP Holdings - BL Direct

Google Scholar Limits

- Limits of Google Scholar have been identified mainly in the lack of clarity about the contents analyzed by the search engine. In fact, no information is given about the sources covered, the type of document processed, or the time span covered
- Apparently, its index includes virtually all peer-reviewed journals available online, except those published by Elsevier. Google Scholar sources include preprint servers as well, although it is not clear how a Web site qualifies for inclusion in its search.

Thomson Web of Science

http://scientific.thomson.com/webofknowledge/wosgenben.html

- Subscription-based multidisciplinary database covering scientific literature from about 6,125 journals which, in some cases, date back to 1945
- Valuable and well-known tool for exhaustive retrieval of research information. Web of Science consists of five databases gathered from thousands of scholarly journals in all areas of research
- A citation index contains the references cited by the authors of the articles covered by the index. The feature "*cited reference*" enables users to find articles that cite a previously published work. In addition to cited reference search, the database can be searched by topic, author, journal title and author address.

Methods

- Two sets of articles published in 2002 were analyzed examining the number of citations retrieved by Google Scholar and Web of Science, and the quality of the results obtained
- The publication year 2002 was selected in order to ensure the possibility of retrieving an adequate number of citations in the subsequent three and a half years.

First Sample

- Selection of the articles published in 2002 by the researchers working in the former Bacteriology and Medical Mycology Laboratory of ISS
- 44 papers, 17 researchers of the laboratory figuring as first authors
- The search for the first set of articles was performed in June 2005 and later in the first ten days of July 2006, both on Web of Science and on Google Scholar, to verify the number of citations received by each article, and to compare results obtained at the distance of a year.

Bibliographic Reference	WoS Citations 2005	WoS Citations 2006	GS Citations 2005	GS Citations 2006
CLIN INFECT DIS 2002 35(2)PP 205-208	5	9	2	7
MICROB DRUG RESIST 2002 8(1) PP 45-53	4	5	0	6
ANTIMICROB AGENTS CHEMOTHER 46(9) 2821-2828	13	16	5	13
CLIN DIAGN LAB IMMUNOL 2002 9(1) PP 66-74	15	19	13	24
J INFECT DIS 2002 185(2) PP 188-195	16	22	9	19
FEMS IMMUNOL MED MICROBIOL 2002 32(3) PP 211-218	4	7	3	5
CELL IMMUNOL 2002 VOL 220(1) PP 30-38	2	3	2	5
ANTIMICR AGEN CHEMOTHER 46(5)1269-1272	24	38	13	35
COMP IMMUNOL MICROB INFECT DIS 2002 25 (4) 217-228	1	1	1	1
MICROB DRUG RESIST 2002 8(1) PP 1-8	9	11	0	8
INFECT IMMUN 2002 VOL 70(12) PP 6621-6627	6	8	4	8
J IMMUNOL 2002 169(11) PP 6231-6235	1	3	0	5
CIRCULATION 2002 106(5) PP 580-584	20	25	13	20
INFECT IMMUN 2002 70(2) PP 985-987	8	10	7	10
MICROB DRUG RESIST 2002 8(2) PP 85-91	5	9	3	9
RES MICROBIOL 2002 153(1) PP 37-44	1	1	1	1
INT J MED MICROBIOL 2002 291(6-7) PP 571-575	8	12	9	12
MICROBIOLOGY 2002 VOL 148 PP 3173-3181	2	2	1	2
MEDICAL MYCOLOGY 2002 40(5) PP 471-478	1	3	1	2
J CLIN MICROBIOL 2002 40(10) PP 3660-3665	5	8	3	6
BLOOD 2002 99(7) PP 2490-2498	17	23	14	24
ANTIMICROB AGENTS CHEMOTHER 2002 46(6) 1688-1694	5	11	4	8
INFECTION IMMUNITY 2002 70(9) PP 4791-4797	6	6	3	7
VACCINE 20(17-18) PP 2229-2239	8	9	4	6
INFECTION IMMUNITY 2002 70(10) PP 5462-5470	19	29	11	27
INFECTION IMMUNITY 2002 70(5) PP 2725-2729	7	11	7	11
MICROBIOLOGY 2002 148 PP 3873-3880	10	12	6	11
J CLIN MICROBIOL 2002 40(3) PP 774-778	37	51	23	54
J CLIN MICROBIOLOGY 2002 40(9) PP 3470-3475	13	21	6	27
J CLIN MICROBIOL 2002 40(7) PP 2662-2665	12	16	9	16
EUR J IMMUNOL 2002 32(11) PP 3050-3058	12	16	8	16
CLIN EXPER IMMUNOL 2002 129(2) PP 254-264	3	3	0	2
J INFECT DISEASES 2002 186(3) PP 351-360	17	25	11	24
TRENDS MOLECUL MEDICINE 2002 8(3) PP 121-126	3	7	2	10
J ANTIMICROBIAL CHEMOTHER 2002 50(1) PP 125-128	2	2	0	3
INTERNAT J TUBERC LUNG DIS 2002 6(1) PP 32-38	8	12	4	12
EUR J CLIN MICROBIOL INFECT DIS 2002 21(3) PP 181-188	7	12	6	10
J CLIN MICROBIOLOGY 2002 40(11) PP 3956-3963	10	18	10	16
J OF MEDICAL MICROBIOL 2002 51(12) PP 1071-1079	1	1	1	2
TRENDS IN MICROBIOLOGY 2002 10(4) PP 177-178	3	3	3	4
MICROBIAL PATHOGENESIS 2002 32(3) PP 135-141	2	2	2	2
J IMMUNOL 2002 169(1) PP 366-374	26	33	22	36
J INFECT DIS 2002 186(1) PP 87-93	4	5	4	10
FEMS MICROBIOL LETTERS 2002 214(1) PP 87-93	9	10	5	10
Total	**391**	**550**	**255**	**546**

Table I. Number of citations retrieved in WoS and in GS for the 1st sample of papers

DATE	WoS Citat	% WoS	GS Citat	% GS	Total	Diff.
June 2005	391	61%	255	39%	646	21%
July 2006	550	50%	546	50%	1096	0%

Results for the first sample

- Although in June 2005 the number of citations retrieved in GS was considerably inferior with respect to those in WoS, the number of citations received by each paper showed a similar trend. The most cited papers were almost always the same in both search tools. Highly cited papers can therefore be easily and reliably extracted from both tools.

Results for the first sample (2)

- In June 2005 the number of citations obtained through WoS was higher (about 20% more) than that obtained through GS
- In 2006 the total number of citations retrieved by the two tools is practically the same
- Results from GS changed dramatically after November 2005. It can be argued that retrieval from GS gives much greater results in 2006. This is probably due to improvements in its search algorithms
- Confirmed by the present study as well.

Second sample of papers

- In the second sample we took into consideration the papers published in 2002 on WoS indexed journals of ten highly-cited Italian authors operating in ISS in very different fields of interest
- Results of the search performed for each author between the dates of July 10th and July 12th 2006 are shown in the following table

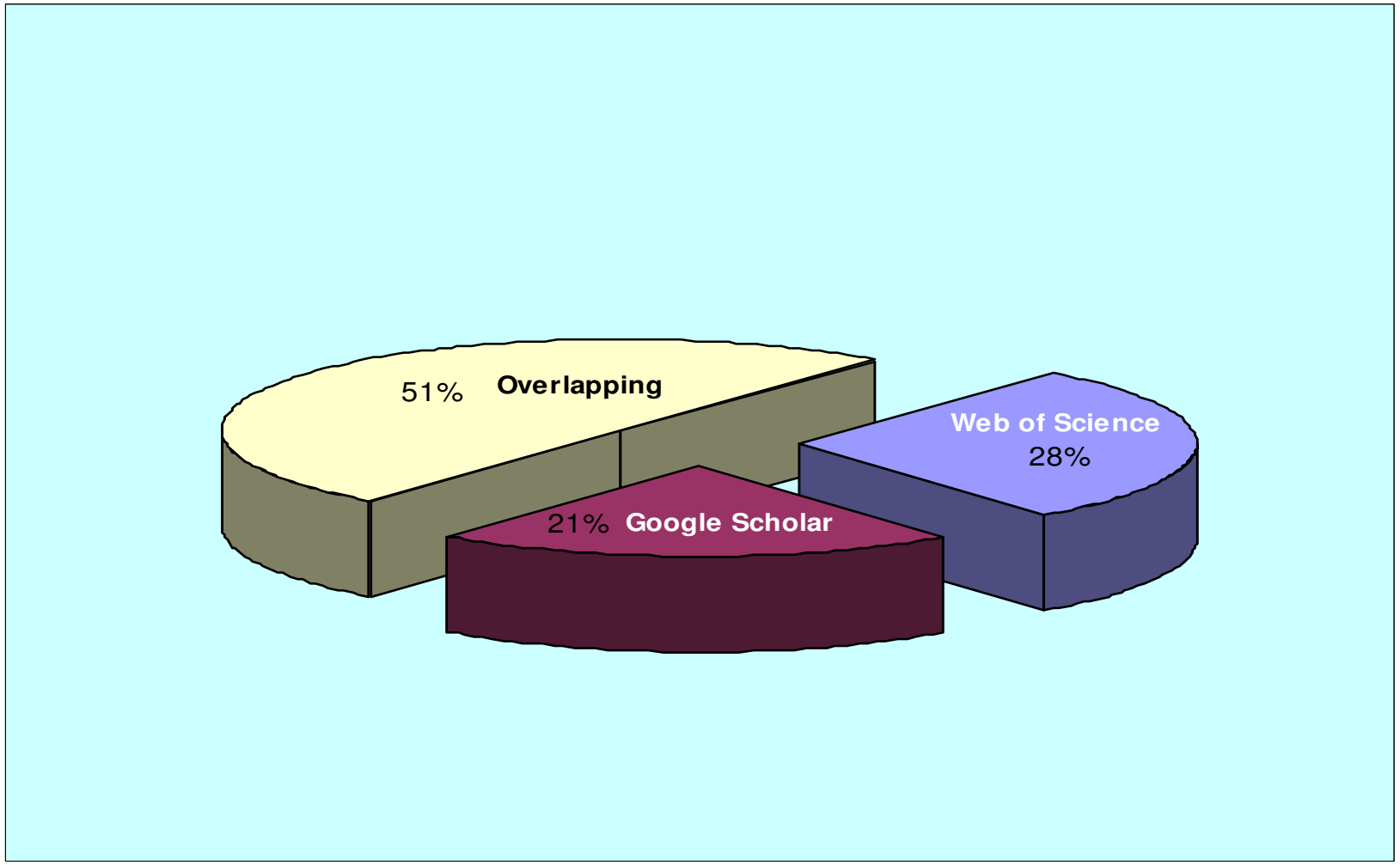
Results for the 2nd sample of papers (Search performed on July 10th -12th 2006)

Department	Papers 2002	WoS	GS	WoS Unique		GS Unique		Overlapping	
	#	#	#	#	%	#	%	#	%
Technology and Health	12	171	160	58	27%	47	22%	113	52%
Cell Biology and Neurosciences	10	301	290	73	20%	62	17%	228	63%
Drug Research and Evaluation	10	73	66	27	29%	20	22%	46	49%
AIDS	16	292	226	130	37%	64	18%	162	46%
Cell Biology and Neurosciences	6	50	46	28	38%	24	32%	22	30%
Infectious, Parasitic, Immune-mediated diseases	13	203	196	69	26%	62	23%	134	51%
Cell Biology and Neurosciences	4	39	42	14	25%	17	30%	25	45%
Haematology, Oncology and Molecular Biology	7	57	46	26	36%	15	21%	31	43%
Food Safety and Veterinary Public Health	10	186	173	59	25%	46	20%	127	55%
Haematology, Oncology and Molecular Biology	9	128	122	56	31%	50	28%	72	40%
Total	97	1500	1367	540	28%	407	21%	960	51%

Results for the second sample

- WoS retrieved about 130 citations more than GS
- After an analysis of the number of unique and overlapping citations in both search tools for the second sample of articles, it was found that WoS retrieved the highest number of unique citations, while the overlap amounted to about 50% of the total material, as shown in the following graphic

Search performed on July 10-12 2006 2nd sample



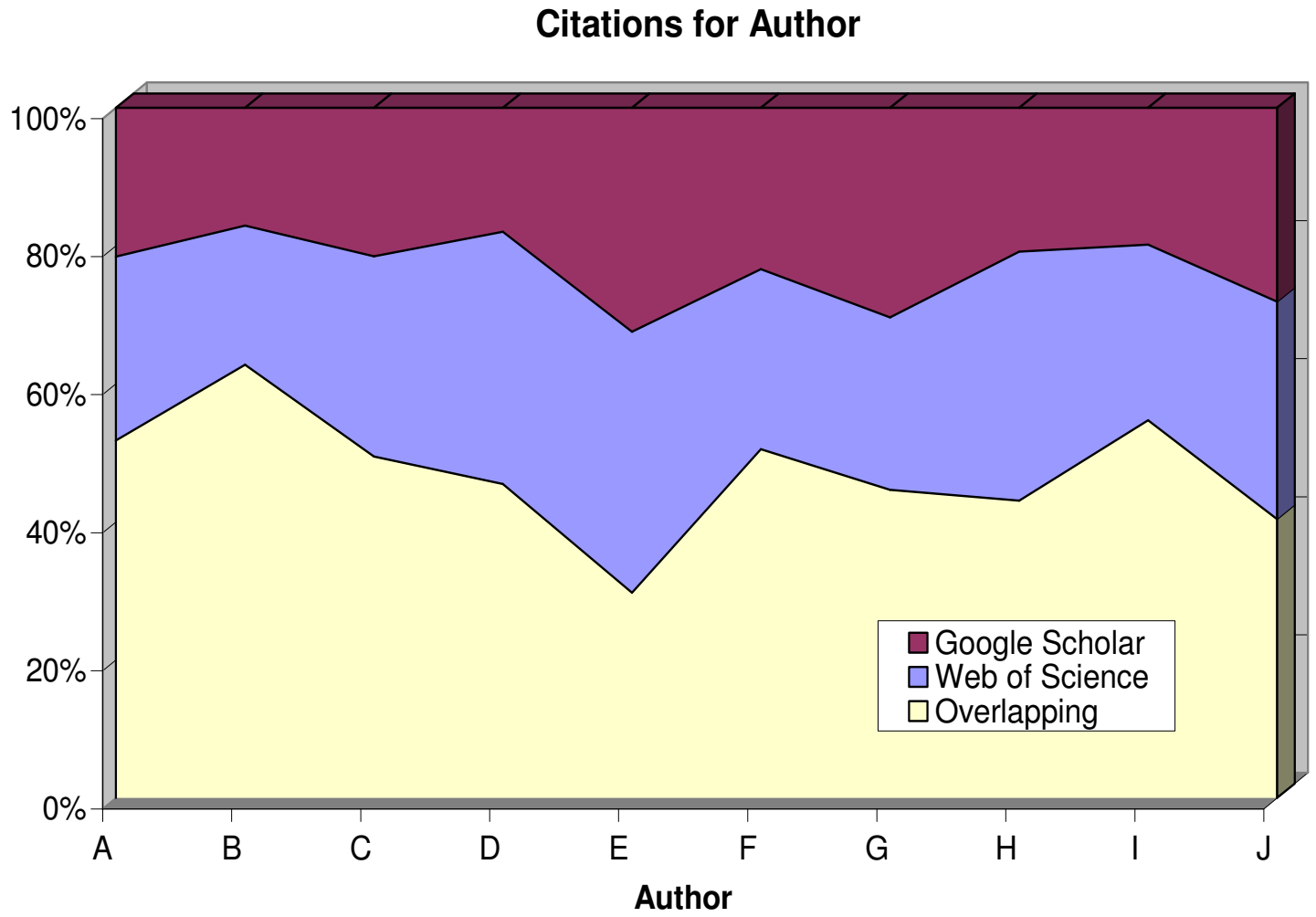
Results from the second sample (2)

- An analysis of the unique citations present in GS reveals that they are extracted primarily from types of documents not present in WoS: technical reports, e-prints, books, briefings, website contents, clinical protocols and journals not indexed by WoS.
- Moreover, GS apparently retrieves many unique citations deriving from documents in languages different from English, such as Chinese.
- The large amount (51%) of overlapping documents confirms the hypothesis of the high quality of the GS search engine for detecting scholarly literature

Unique and overlapping citations

- In the following graphic the trend of unique and overlapping citations for each author is presented
- The trend is similar for each author, without regard to the different disciplines, and consequently also for the different journals in which their articles were published.

Percentage of unique and overlapping citations for each author of the 2nd sample in Google Scholar and in Web of Science



Conclusions

- This study analyzed the amount of citations received for two defined sets of scientific articles published in 2002
- The first set was made up of 44 papers, while the second one included 97
- Although the numeric size of the sample was not high, some interesting results could be extracted from available data
- Based on this preliminary analysis, and in consideration of the improvement of its technology in recent months, Google Scholar can be considered a quality source for citation retrieval. Nevertheless, a comprehensive retrieval can be achieved only through the use of quite expensive subscription-based tools, such as Web of Science.

Conclusions (2)

- However, for those libraries or institutions which cannot afford subscription costs, Google Scholar can be considered a valuable source of information for citation tracking.
- Moreover, since resulting citations do not match exactly, due to the difference in type of documentation considered by search algorithms, GS can still be used to integrate findings from Web of Science
- The two instruments can be seen as complementary, although a good percentage of overlapping material results from this research.