SOME OF THE MOST CITED RESEARCH IN THE LAST CENTURY

Maria Daniela Bondoc[1], Claudia Burtescu[2]

[1] University of Pitesti, e-mail: daniela.bondoc@upit.ro
[2] University of Pitesti, e-mail: claudia.burtescu@upit.ro

Abstract. The content of this article describes and analyzes two famous lists of the most prestigious research papers of the last century starting from databases and citations of WoS or Thomson Reuters, which in the meantime has become Clarivate Analytics and Google Academic (Google Scholar). If the first list is dominated by articles, in the second one books have an important place. Two findings are relevant: the continuity over time of the biology dominance together with the other bio-sciences (Biology lab technique and Bioinformatics) over a period of more than half a century, and the tendency to increase the article quotations in relation to the books.

Keywords: citations, Google Scholar, Web of Science, research

1. INTRODUCTION

Three questions can best open this work:

a) What are the most quoted articles and books in the last century?
b) What are the research areas most often quoted as references and cited with the greatest prospects in modern trans-, inter-, cros- and multidisciplinary research?
c) Does the article or book become the most important source of quotation and research in the modern list of internal or final final references?

There are differences between mass research (Google scholar or Google academic done by professors, students, master students etc) and the elite research (ISI magazines indexed in the Web of Science, currently Clarivate Analytics), such that starting from an article published in october 2014 in Nature magazine [1], we aim to highlight the most quoted scientific papers in the Web of Science (WoS) and Google Scholar (S) databases.

The number of quotations is considered "a proper indicator of the quality and relevance of the article as well as of the scientific level of the researcher's work productivity. The higher the interest in a particular research problem, the more the number of those who quote it as a relevant source of information is increasing" [2].

A large number of quotations is a confirmation that the work has an impact on the scientific or the academic community.

Our study analyzes the list of the first 100 works quoted in the Web of Science (WoS) [3] and Google Scholar (GS) [4], available from the Nature magazine website.

An investigation of great interest is related to the scientific field of belonging of the most quoted 100 papers according to the lists (WoS and GS)

2. WEB OF SCIENCE / CLARIVATE ANALYTICS LIST

Web of Science Core Collections has indexed 11.9 million documents from 251 scientific categories from 2010 to 2014 [5]. This list includes mainly research in the fields of pure sciences, engineering and medicine, while humanities or social sciences are less to be found.

The analysis of this list has highlighted, first of all the structure of the most quoted papers by field. The WoS list with the most quoted 101 papers comprises 10 domains, of which Biology lab technique is distinguished by 39 papers, followed by Physical chemistry with 16 papers, as can be seen in Table 1.

Table 1 Domain break down of the most quoted papers in the WoS list

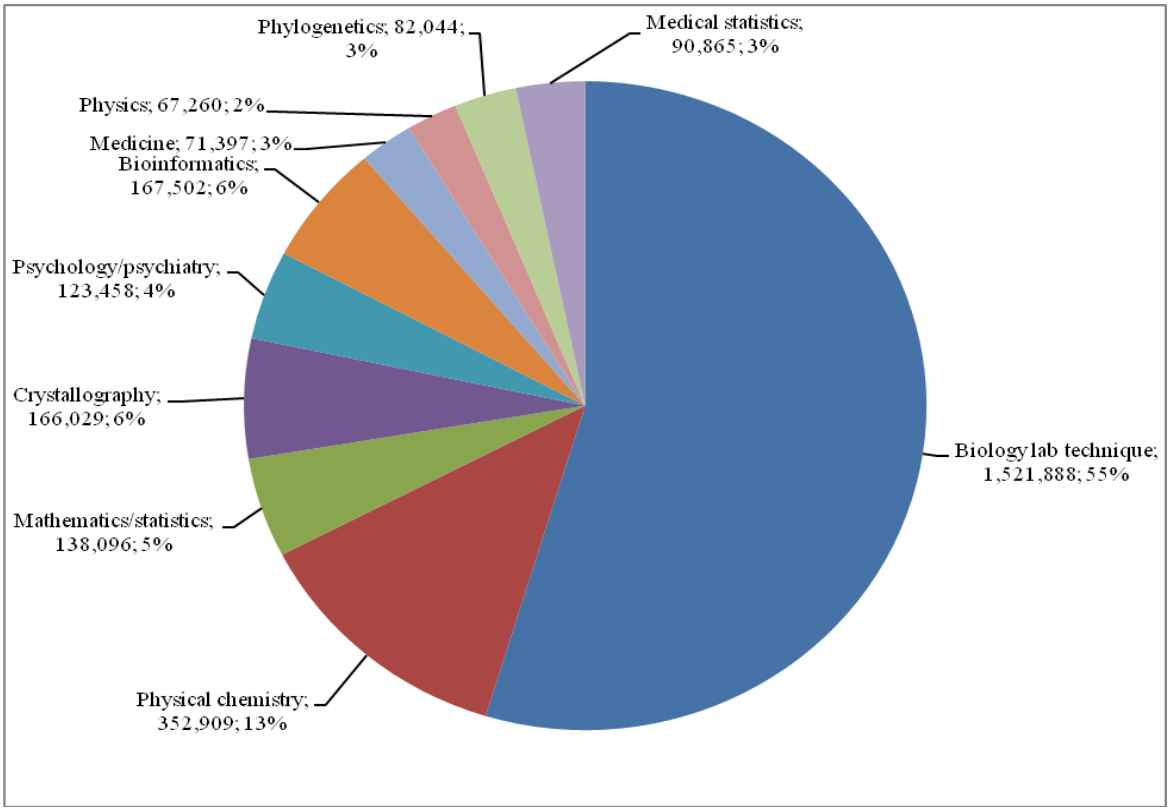
Subject	Sum of times cited	Count of subject	Best ranked of each subject		
			Times cited	Title	Year
Biology lab technique	1,521,888	39	305,148	*Protein measurement with the folin phenol reagent.*	1951
Physical chemistry	352,909	16	46,702	*Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.*	1988
Mathematics/ statistics	138,096	9	18,203	*Fuzzy sets.*	1965
Crystallography	166,029	9	37,978	*A short history of SHELX.*	2008
Psychology/ psychiatry	123,458	6	34,532	*"Mini-mental state": A practical method for grading cognitive state of patients for clinician.*	1975
Bioinformatics	167,502	6	40,289	*Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.*	1994

Subject	Sum of times cited	Count of subject	Best ranked of each subject		
			Times cited	Title	Year
Medicine	71,397	5	17,220	*Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease.*	1984
Physics	67,260	4	22,899	*Helical microtubules of graphitic carbon.*	1991
Phylogenetics	82,044	4	30,176	*The neighbor-joining method: A new method for reconstructing phylogenetic trees.*	1987
Medical statistics	90,865	3	38,600	*Nonparametric estimation from incomplete observations.*	1958

The total number of quotations of the Top 101 WoS papers is 2,781,448 of which 1,521,888 are for the Biology lab technique which is best represented in the WoS list and in terms of the number of quotations, 55%, followed by Physical chemistry, with 13%.

The result of the distribution of the number of quotations in the 10 fields is shown below.

Fig. 1 Field distribution of quotations from the WoS list



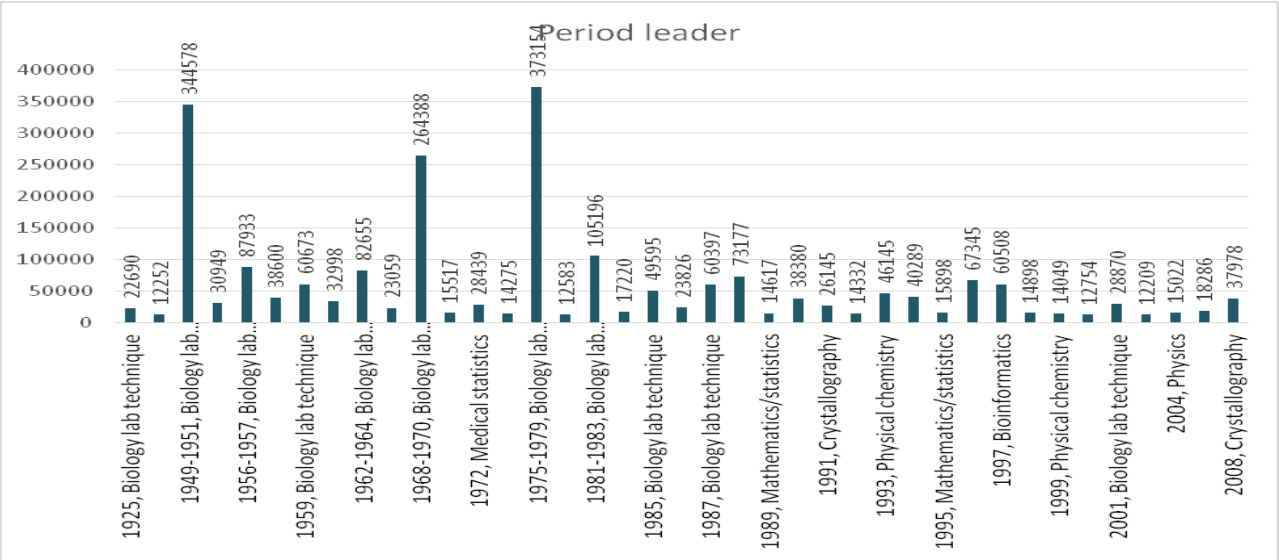
Source: realised by authors from WebofSciencetop100.xls and Google Scholar Top 100.xls [online]

From Fig. 1 the dominant structure (55%) is represented by papers in biology and derived biosciences (including laboratory biology or specific techniques) according to the WoS list. The paper with the highest number of quotations (305,148) in WoS is Protein measurement with the folin phenol reagent, published in 1951 by the Journal of Biological Chemistry, owned and published by the American Society for Biochemistry and Molecular Biology, Inc.

Figure 2 shows that over a period of over 80 years (1925 – 2005) Biology together with the other biosciences (Biology lab technique and Bioinformatics) chronologically dominate the volume of quotations with a number of 54 years.

On the second place, until the end of the 20th century, we have Mathematics / Statistics followed by Physical chemistry (including Physics).

Fig. 2 Analysis of the quotations dominant subject between 1925 - 2005



Source: realized by authors after WebofSciencetop100.xls and Google Scholar Top 100.xls [online]

3. GOOGLE SCHOLAR LIST

Google Scholar is "a free online academic search engine that uses automated software to extract citations from on-line digital publications and combines it with data provided by some publishers". [6]

Other authors [7] emphasize the fact that „the free availability of GS allows for a democratization of citation analysis as it provides every academic with access to citation data regardless of their institution's financial means".

The Google Scholar List (Google Academic) of the top 100 most quoted research includes 64 books and 36 papers. The analysis of papers compared to the structure of the quotations revelas some relatively contradictory aspects from the structural and chronological point of view, especially on the long term. Thus, in figure no. 3 it is found that in the first 100 quoted research the books dominate with a percentage of 64%, but in the number of quotations their structural dominance decreases to 59%.

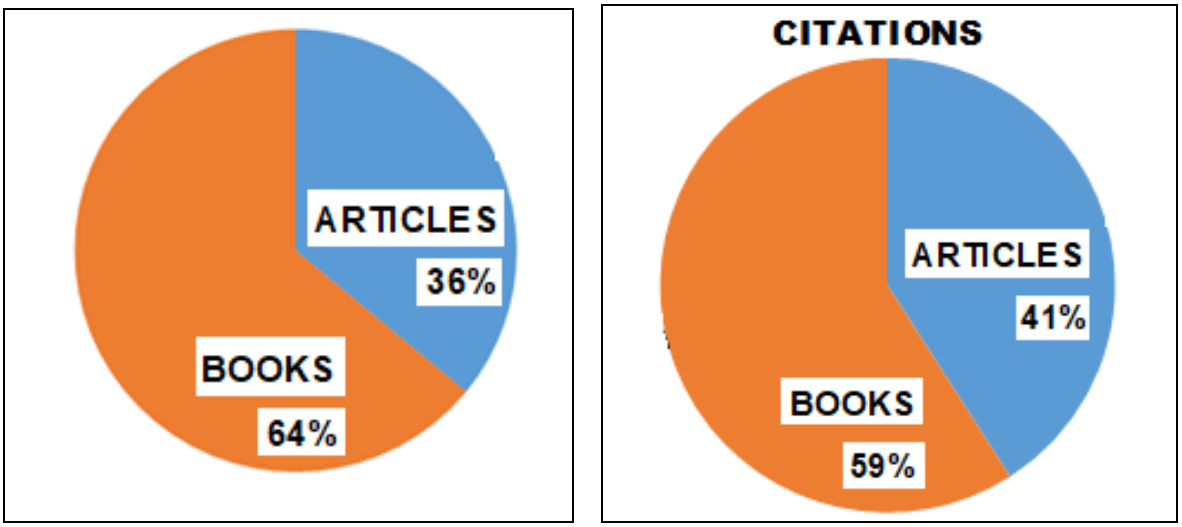


Fig. 3 Emphasizing the position of articles in the confrontation between the number of papers and quotations

Table 2 lists the oldest 5 books or articles found among the top 100 most quoted papers of GS.

Table 2 Top 5 oldest papers in GS

Rank	Authors	Title	Year	Number of Citations	Book / Article
80	Smith, A. & Nicholson, J. S.	*An inquiry into the nature and causes of the Wealth of Nations*	1887	33,435	Book
9	Shannon, C. E.	*A mathematical theory of communication*	1948	69,273	Article
52	Feller, W.	*An introduction to probability theory and its applications*	1950	42,290	Book
2	Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J.	*Protein measurement with the folin phenol reagent.*	1951	192,710	Article
34	Greene, W. H.	*Econometric analysis*	1951	48,100	Book

The article with the highest number of quotations in Google Scholar (223,131) is Cleavage of structural proteins during the assembly of the head of bacteriophage T4, published in 1970 by the prestigious Nature magazine. This is followed by Protein measurement with the folin phenol reagent, with 192,710 quotations (the same article we have found in the top Web of Science list). In the top 100 of the most quoted 100 papers in Google Scholar there is a book from 1887, authored by the well known Scottish economist Adam Smith and Joseph Shield Nicholson, *An inquiry into the nature and causes of the Wealth of Nations*).

We note the presence of a book of *Econometric analysis* in this list, that was published later (the 8th edition in 2017).

4. A "CONFRONTATION" OF THE TWO LISTS (WoS / GS) PER INTERVALS OF TIME

We also took into consideration the emphasis of the most quoted papers according to the time of their publication, spred over time periods. The data in Table 3 together with Figures 4 and 5 illustrate this aspect.

Table 3 Number of papers from the top of the two lists, spread on time intervals

Time interval	Number of research in WoS top	Number of research in GS top
Before 1970	25	28
Between 1971 - 1990	52	48
Between 1991 - 2000	20	21
Between 2001 - 2010	3	3
After 2011	0	0

Regardless of the top 100 most quoted papers (Web of Science or Google Scholar), their dominant structure belongs to the period between 1971 – 1990, although it was expected that the larger and older period 1887 – 1970 would have a significantly higher share, after which the quote speed decreases decade after decade with all the improvements brought by modern technology in the system of quotations and references on the Internet.

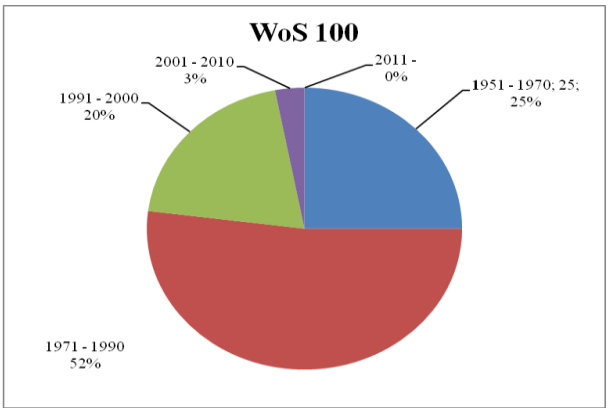


Fig. 4 Structure by time periods of articles in the WoS top

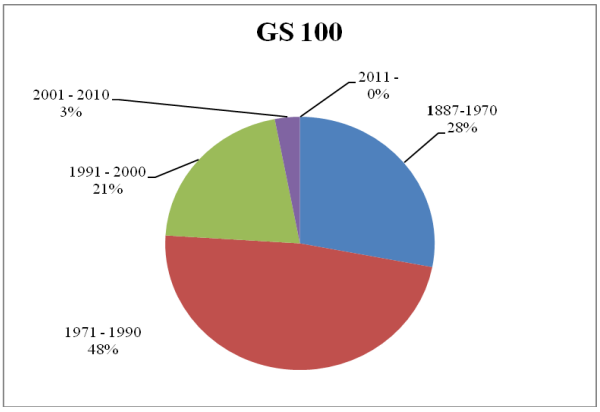


Fig. 5 Structure by time periods of articles in the GS top

In both lists, the period 1971 – 1990 has the main weight, while the recent research (after 2000) are being poorly represented.

5. CONCLUSIONS

In both WoS and GS tops, the first three papers based on the quotations number are represented by articles in the field of biology research, the only ones with over 100.000 quotations (the first one having over 300.000 quotations in WoS) published in 1951, 1972 and 1976.

The first positions in the two lists are occupied by the same paper, but reversed: the first ranked in WoS (Protein measurement with the folin phenol reagent) is found in the second position in the GS rankings while the paper with the highest number of quotations in GS (Cleavage of structural proteins during the assembly of the head of bacteriophage T4) is the second in the WoS list.

What's next? We propose further in – depth studies (Data mining) for:

- highlighting the words most commonly found in the WOS100 and GS100 titles
- highlighting the weight of fundamental scientific areas and new areas
- highlighting geographic areas (GIS) of authors vs. Publications (magazines / publishers)

The two lists of the most quoted research papers of the last century (Web of Science and Google Scholar) highlighted some distinct and relatively surprising aspects.

6. REFERENCES

[1] Van Noorden, R., Maher, B., And Nuzzo, R. (2014). *The Top 100 Papers Nature explores the most-cited research of all time*, Nature, vol. 514, pp. 550-553. [online] Available at: http://www.nature.com/news/the-top-100-papers-1.16224 [Accessed November 23, 2017].

[2] Dinu, V., Săvoiu, Gh, Dabija, D.-C. (2016). *Conceiving, Writing and Publishing a Scientific Paper - An Approach in the Context of Economic Research* [A concepe, a redacta și a publica un articol științific. O abordare în contextul cercetării economice]. Bucharest, Romania: Editura ASE, 131.

[3] WebofSciencetop100.xls [online] Available at: www.nature.com [Accessed November 20, 2017].

[4] Google Scholar Top 100.xls [online] Available at: www.nature.com [Accessed November 20, 2017].

[5] Patience, G.S., Patience, C.A, Blais, B., Bertrand, F. (2017). *Citation analysis of scientific categories*. Heliyon 3 (5), e00300, May 2017. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5446096/pdf/main.pdf [Accessed November 21, 2017].

[6] Thelwall, M; Kousha, K. (2015). *Web Indicators for Research Evaluation. Part 1: Citations and Links to Academic Articles from the Web*. El profesional de la información, 2015, septiembre-octubre, v. 24, n. 5. eISSN: 1699-240. [online] Available at: https://www.researchgate.net/profile/Kayvan_Kousha/publication/282613558_Web_indicators_for_research_evaluation_Part_1_Citations_and_links_to_academic_articles_from_the_Web/links/561430a508aec62244100f09/Web-indicators-for-research-evaluation-Part-1-Citations-and-links-to-academic-articles-from-the-Web.pdf [Accessed November 28, 2017].

[7] Harzing AWK, Van der Wai R. *Google Scholar as a new source for citation analysis*. Ethics in Science and Environmental Politics (ESEP). 2008; 8(1):61–73. [online] Available at: http://www.int-res.com/articles/esep2008/8/e008p061.pdf [Accessed November 28, 2017].