

Bibliometric analysis at the National Technical University Athens: the case of the faculty publications

Alexandros Iliakis, Christina Anastasopoulou

Central Library, National Technical University of Athens, Greece.

Abstract: Within the framework of the internal evaluation of the National Technical University of Athens, as specified by the standards of the Hellenic Quality Assurance and Certification Authority in Higher Education, the members of the Library of the National Technical University of Athens were asked to provide information on the publishing activity of the members of its Academic Staff. After consulting with the Management and the Quality Assurance Unit of the National Technical University of Athens, it was agreed to implement a bibliometric analysis of the publications - mainly in international scientific journals - of its academic staff. In particular, it was suggested to use scientific publications - up to and including 2016 - for each active member of the academic staff. The information that had to be collected was the following: the number of publications, the number of references and the h-index.

Keywords: Bibliometric analysis, research analysis, publications, teaching staff, methodology Scopus, Publish or Perish, National Technical University of Athens.

1. Introduction

The Hellenic Quality Assurance and Accreditation Agency (HQA) in Higher Education was created in accordance with the Legislation (Laws No. 4115/2013, 4076/2012, 4009/2011, 3374/2005), aiming at the formulation and implementation of a single quality assurance system as a reference system for achievements and the work carried out by Higher Education Institutions. Another goal of the HQA is to gather and encode critical information that will guide the Government in effectively enhancing Higher Education.

The role of the Quality Assurance Unit (MODIP) of the Greek Higher Education Institutions (Law No. 4009/2011, Article 14) is summarized in the following four rules: (a) development of policies, strategies and processes to

improve the Institution, (b) Organization, function and improvement of the quality assurance system, (c) Coordination and support of the evaluation processes of units and (d) Support of the processes for external evaluation and certification of the Programs of Studies.

Each institution is responsible for ensuring and continuously improving the quality of its educational and research output. Additionally for the efficient operation and performance of its services, in line with international practices, in particular those of the European Higher Education Area, and the principles and guidelines of the HQA. For the above purpose responsible at each Higher Education Institution (HEI) is the Quality Assurance Unit (MODIP).

2. Evaluation of Greek Higher Education Institutions

The evaluation of higher education institutions takes place in two stages:

- i. The first stage concerns the evaluation of the work carried out by the higher education institutions themselves in relation to their physiognomy, objectives and mission (internal evaluation). Thus, the individual academic units of each higher education institution submit each year to the MODIP an internal report. Based on them the MODIP draws up the institution's internal evaluation every two years.
- ii. The second step is the external evaluation of the work being carried out by a committee of independent experts, which also takes into account the results of the internal evaluation report of the academic unit.

Templates of forms, which are provided by the HQA, are used in both stages.

For the National Technical University of Athens (NTUA) the Quality Assurance Unit (MODIP-NTUA) undertook the evaluation process under the supervision of the Vice-Rector of Academic Affairs. During the preparation of the internal evaluation report, the library was asked to assist in the collection of the required data: the publications of the Teaching and Scientific Personnel (TSP), as identified in the tables of the template of the HQA. It included the number of scientific publications by the TSP of the NTUA and the influence of their research.

The Schools of the NTUA had gathered the requested information and were included in their internal evaluation reports. The administration felt it advisable to carry out a concentrated effort to collect more accurate data and sub-indices that were not included in the reporting standards.

3. Bibliometric analysis of the publications of the TSP of the NTUA.

In the spring of 2016, the Vice-Rector of Academic Affairs asked library staff for information on the publications of the TSP of the NTUA.

To gather accurate information about these publications it was decided to use bibliometric methods.

The measurement of the influence of a scientist-author is called bibliometrics. The technique for the verification of this influence or effect ranges from simple measurements of scientist publications to sophisticated mathematical equations. Two of the most well-known bibliometrics methods are: the Impact Factor commonly applied in journals, and the h-index, which is usually applied to authors.

The Impact Factor, proposed by Eugene Garfield, is the measure between citations and recent articles published. Thus, the Impact Factor of a journal is calculated by dividing the number of peer-reviewed journals published in the two previous years by the number of published articles in that period. Journal Citation Reports calculate and publish their annuals coefficients impact factor for a number of journals. A higher impact factor generally indicates that the articles of this journal have been cited more.

The h-index was proposed by Jorge Hirsch in 2005 as an alternative to the Impact Factor. The h-index quantifies scientific production and influence of a scientist based on the articles cited more and the number of citations that the publications have received from other scientists. For example, a author of a journal with h-index 30 has written at least 30 articles, each with at least 30 citations. Thus, a higher h-index indicates the most frequently cited publications. This indicator is useful because it takes into account the heterogeneous weight of the articles between those that have been extensively cited and those that have not yet been cited.

Thus it was decided to apply these two indicators (Impact Factor and h-index) for the evaluation of the academic work of the TSP of the NTUA. In order to better explain the potential of these indicators, searches were made for the professors who were members of the Rectorate in Scopus and Web of Science and were presented the results.

In a further investigation, it was argued that there was a previous effort of collecting the publications of the TSP of the NTUA. Specifically, for the enrichment of the institutional repository a team of the Hellenic Academic Libraries Link (HEAL-Link), designed a software application to collect the publications retrospectively until 2013 from Scopus, Web of Science and Microsoft Academic. This action did not include the collection of any bibliometric data.

In the discussions that followed with the Deputy Rector of Academic Affairs and the colleague of the MODIP-NTUA. it was decided to collect the h-index, the number of publications and the citations analysis each member of the TSP of the NTUA. They were informed by the Library of the various applications that can measure the impact of research and of the publications at the institution's level. Two platforms were proposed, [SciVal](#) and the [InCites](#). These applications

have been presented in the past to HEAL-Link and the NTUA. These specific proposals were rejected due to insufficient funds. Alternative measurement methods, such as Altmetrics were suggested but there were also rejected because of their cost.

Ultimately, the only remaining option was to perform manual searches on selected databases: a) Google Scholar as it is open access and b) the Scopus database. The second was selected because there was an active subscription through HEAL-Link, was considered to cover better the engineering sciences and the author search function that shows their h-index and citation analysis. The Google Scholar itself did not provide bibliometric data so a free application was selected: Publish or Perish by Ann-Wil Hersing. A workgroup was set up to collect the data with staff from the Central Library and HEAL-Link.

The Directorate of Teaching Personnel, at the request of the library, provided a list of active members of the TSP with the following data: surname, first name, father's name, school and grade. This list was uploaded on a Google worksheet and the names were divided among the workgroup. This practice was preferred because all members had access to a file which would be processed at the same time. A sample of the completed table follows:

Table 1. Details of the TSP of NTUA

Grade	Last name	Name	School
3. EPIKOUROS	Spyropoulou	Ioanna	SCHOOL RURAL AND SYRVEYING ENGINEERING
2.FULL PROFESSOR	Chaniotou	Helen	SCHOOL OF ARCHITECTURE

To upload the results from the searches in Publish or Perish and in the Scopus database the following columns were added to the worksheet.

Table 2. Data from Google Scholar and Scopus

h-index	Number of publications	Citations	h-index Scopus	Number of Scopus publications	Citations Scopus
11	106	425	8	61	184
0	2	0	-	1	0

The period that the research would cover was an important issue because at the beginning the MODIP-NTUA requested data only for 2016. In the discussions, the Library team argued that according to international practice the research

covers a range of five years. So it was decided that it would run for the period from 2011 to 2016.

When searching for the names, certain issues arose with the transliteration of the names of the TSP in English. For this reason, we had to search by every possible combination. To facilitate workflow and for future use, additional columns were added.

Table 3. Revised data

Name (English)	Middle name	Surname (English)	e-mail
Milt / M. / Miltos / Miltiades	KC	Statheropoulos	stath@ntua.gr
Elias / Ilias		Zacharopoulos	zachar@ntua.gr
Ioannis / Yiannis	A	Polyrakis	polyr@ntua.gr
Ioannis / J / Ioannis	K.	Chatzigeorgiou	chatz@ntua.gr

During the research an issue came up in regard to the members of the TSP, who had left or retired from the NTUA during the period from 2011 to 2016. Some names were included in the list but the majority of those who had retired or left, were absent. The decision was to be included in the research but to end in the year of departure. The Directorate of Teaching Personnel provided the names and they were added to the worksheet with additional columns to separate them from the active members.

Table 4. Retired TSP

Grade	Last name	Name	School	Retired	YearOfRet
TEACHERS OF FULL EMPLOYMENT	PAUTZHI	MARY	SCHOOL OF RURAL AND SURVEYING ENGINEERING	1	2011
TEACHERS OF FULL EMPLOYMENT	ROKKOS	CONSTANTINE	SCHOOL OF RURAL AND SURVEYING ENGINEERING	1	2012

In Publish or Perish, the search besides the time limit (2011-2016) required more restrictions due to the presence of common names. For this purpose, this research was further limited by the use of the affiliation and the country. While in many cases it was necessary to check the publications and those that differed from the teacher's subject matter, were removed from the calculation of the indicators.

The search mode in Scopus is different. The first time an author publishes an article, a code is assigned to him which is called Scopus Author id: So we researched based on the name of the author but this time to locate the profile that is created from Scopus. In the case of TSP of the NTUA several had multiple Authors ids due to variations of their name or of the institution with which they were affiliated. During the manual search phase, it was impossible to combine the results from different profiles so we selected the one associated with the NTUA or the one with the most data.

In any case, manual search required too much time and then it would take more time to integrate the results, process and export the statistics. We have to note that the TSP of the School of Architecture was found to have low indicators. This is because they make few publications in journals and show their work in exhibitions and give lectures.

The workgroup was informed by a professor at the School of Chemical Engineering that within the framework of their internal evaluation an application was developed to automatically retrieve the publications and corresponding citations from the Scopus database. The team approached Professor Kyranoudis, who made a presentation of how the software worked and provided it for testing.

The application used the Scopus API for collecting the data based on the Scopus Author id. Since the manual searches were already made on the database, we needed to collect the Scopus Author id's so additional columns were created.

Table 5 . Scopus Author ID

Author Id to search	Author id NTUA (comma separated)	Author id (Not NTUA)
7006736240 + OR + 56566276900 + OR + 7006736235 + OR + 7006736241	7006736240, 56566276900, 7006736235	7006736241
56765618300 + OR + 34976139200 + OR + 7004676843	56765618300, 34976139200	7004676843

Various tests of the application were performed by the HEAL-Link IT staff and modified it to work at institutional level and to store all the requested information in a database. To manage the data the following lists were created: a) of the authors, with their details in the NTUA (Authors), b) of their publications (Documents), c) of the citations of the publications (Citations) and d) a list for Authors, Documents, Citations with the relationships between them.

The production of the requested data was done by compiling and executing specific queries in the form of SQL queries per School and in total for the NTUA. The final results were presented on worksheets.

The application has been customized so that articles can be deduplicated. In particular, joint publications by members of the TSP of the same School were counted on the data of each author, but only once in the results of the school. A similar policy was followed at institutional level, and the joint publications by members of the TSP of different Schools were counted at the level of the author, at the school level but once at the institution level.

Note that the following issues have been identified that affect the accuracy of the results:

- The author's manual identification of their Scopus author id.
- Low coverage of the School of Architecture, as well as in the manual research method, they do not publish in journals but participate in exhibitions and give lectures.
- Multiple Scopus author ids (different partner institutions - different name inscription).
- Duplicate publications at the author's level. To solve it there was a deduplication process based on the Scopus Document-ID and / or the DOI (Digital Object Identifier).
- There is no coverage of all publications, since Scopus indexes specific titles at specific periods.
- It is not possible to locate publications in Greek sources.

4. Conclusions

The effort to collect bibliometric data for the publications lasted for 3 months. This process had no cost for the University because it involved permanent employees and contract staff who were already employed in projects of the HEAL-Link. As mentioned above, commercial solutions were almost immediately rejected due to a lack of funds.

Manual search with the Publish or Perish and within the Scopus database, employed 6 people on a daily basis, while the final data processing and statistical would require much time and extra workload. Note that at the time of data collection, these staff abstained from their scheduled work increasing the burden of other staff at the Central Library.

From the experience gained, the team concluded that the use of an automated method of collecting data is a one-way solution. The comparison of the two data sets (manual and automatic) showed the accuracy of the application's operation because it was more complete and provided processing capabilities that were difficult to make in the data that were collected manually. A typical example was the deduplication of the publications at the school and institution level.

The results of the application were accepted by the Rectorate and the MODIP-NTUA, as opposed to the manual ones, because they were very different. The team found that the test code and the database was successful without overseeing that it was a momentary image with important parameters that need to be investigated.

We proposed to the Rectorate the development of the application further to draw on additional data, such as keywords, co-authors and the Impact Factor of journals and of articles. The proposal also included the development of an interface so that it would be used by the staff of the library and the MODIP-NTUA without the need for continuous engagement of a technician. The project was suspended due to the departure of the IT personnel of the HEAL-Link team. The library is currently collaborating with the Directorate of IT of the University to run the application as requested by the Rectorate and the MODIP-NTUA.

References

- Ball, P. (2005). Index is for fair ranking of scientists. *Nature* 436 (7053), 900. <http://dx.doi.org/10.1038/436900a>
- Bornmann, L, and Hans-Dieter D. (2005). Does the h-index for ranking scientists really work? *Scientometrics* Vol. 65, No. 3, 391-392. <http://dx.doi.org/10.1007/s11192-005-0281-4>
- Bornmann, L. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology* Vol. 58, No. 9, 1381-1385. <https://doi.org/10.1002/asi.20609>
- Bornmann, L, Rdiger, M, and Hans-Dieter, D. (2008). Are there better indices for evaluation purposes than the h index?: a comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, 5, 830-837. <http://www3.interscience.wiley.com/cgi-bin/fulltext/117908948/HTMLSTART>
- Costas, R, and Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics* Vol. 1, No. 3, 193-203. <http://www.sciencedirect.com/science/article/B83WV-4NCK2JT-1/1/ddf6c25f8810c462fb8c22ae3d5b5d28>
- Craig, D., Plume, A. et al. (2007). Do open access articles have greater citation impact? a critical review of the literature. *Journal of Informetrics* Vol. 1, No. 3, 239-

248. <http://www.sciencedirect.com/science/article/B83WV-4P18BNV-1/1/0d60463e068ba604e67030229ce6d33b>

Garfield, E. (1999). Journal impact factor: a brief review. *CMAJ* Vol. 161, No. 8, 979-80. <http://www.cmaj.ca/cgi/content/full/161/8/979>

Hirsch, JE. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* Vol. 102, No. 46, <http://www.pnas.org/content/102/46/16569.abstract>

Kouis D, Kalavrouziotis V. &Veranis, G. (2016). Bibliometric data calculation for the evaluation of the academic work of an institution, Workshop presentation session at the *European Library Automation Group Conference* held in Athens, Greece.

Leydesdorff, L (2008). Caveats for the use of citation indicators in research and journal evaluations. *Journal of the American Society for Information Science and Technology*, Vol. 89, No. 2, 278-287. <http://www3.interscience.wiley.com/cgi-bin/fulltext/116839393/HTMLSTART>

Lipsky, P.E. (2007). The lack of impact of the 'impact factor' in clinical medicine. *Nature Clinical Practice Rheumatology*, Vol. 3, No. 4, 189. <http://dx.doi.org/10.1038/ncprheum0462>

Molinari, J, and Molinari A. (2008). A new methodology for ranking scientific institutions. *Scientometrics*, Vo. 75, No. 1, 163-174. <http://dx.doi.org/10.1007/s11192-007-1853-2>

PLoS Medicine Editors (2006). The impact factor game. It is time to find a better way to evaluate the scientific literature. *PLoS Medicine*, Vo.3, No. 6. <http://dx.doi.org/10.1371/journal.pmed.0030291>

Schreiber, M. (2008). An empirical investigation of the *g*-index for 26 physicists in comparison with the *h*-index, the *A*-index, and the *R*-index. *Journal of the American Society for Information Science and Technology* 59 (9): 1513-1522. <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20856>

Thomson Reuters.(2008). Using bibliometrics: a guide to evaluating research performance with citation data [white paper]. <http://scientific.thomsonreuters.com/news/newsletter/2008-07/8465001/>

Wikipedia contributors.(2008). H-index. *Wikipedia, the Free Encyclopedia*, <http://en.wikipedia.org/wiki/H-index>