Article DOI: 10.22032/dbt.39328 Proceedings DOI: 10.22032/dbt.39296

Bibliometric analysis of the publication output of TU Ilmenau in the period 2012-2016

Bernd Markscheffel, Ariane Hertlein and Shihao Zhao

Bernd.Markscheffel|Ariane.Hertlein|Shihao.Zhao@tu-ilmenau.de
Institut für Wirtschaftsinformatik (Business Information Systems Engineering), Department of Economic Sciences and Media, Technische Universität Ilmenau, 98684 Ilmenau, Germany

Abstract

Bibliometrics enables quantitative analysis of research achievements by applying mathematical methods to publishing behaviour whose results can be compared within certain limits. This paper uses a bibliometric analysis to examine the publication performance of scientists at the Technical University of Ilmenau. The publications in the Web of Science for the period 2012-2016 is used as the data basis. With the help of various indicators, an objective picture of the research activities in the above-mentioned period is attempted to be drawn. The results of the bibliometric analysis are used for comparisons at different levels and then presented in different rankings. All collected data are stored in a data base structured form so that it is available as a starting point for future investigations.

Introduction

The number of scientific publications is increasing exponentially, whereas the receptiveness of the individual scientists remains limited. At the same time, there is a desire to evaluate scientific achievements, for example to gain an overview of research, to recognize trends in advance and to be able to assess the efficiency of research (Tunger, 2013).

At the beginning of the 20th century, the first bibliometric analyses were carried out, which made it possible to make statements about the quantity of the various publications (Nix, 2010). These analyses make it possible to evaluate a research achievement, among other things by citations and the number of already published own publications, whereby with increasing citation rate the supposedly most important contributions within a research area can be made recognizable. This in turn can be an impulse for other scientists to take a closer look at these publications (Tunger, 2013).

Due to rising research costs, it is nowadays necessary to present oneself as an institution with strong publishing and reputation in order to a) increase one's visibility and b) effectively raise additional research funds (third-party funds) (Havemann, 2009). Furthermore, there are ranking procedures for educational institutions, such as the CWTS Leiden Ranking (CTWS, 2017) or the Shanghai Ranking (Shanghai, 2017), in which the world's most influential universities are measured by publication output, among other things.

It is therefore becoming increasingly important to monitor the quantitative research performance of one's own institution at different levels of granularity (subject area, institute, faculty) on the basis of publication output in order to be able to assess status and competitiveness and to identify conclusions for potential expansion possibilities in research (Nix, 2010).

Methodology and Indicators

For a basic understanding of bibliometric analysis, a literature review according to Webster & Watson (2002) is carried out. This includes the collection of basic information on bibliometrics, the delimitation of terms and potential indicators. The following search terms are used: *Bibliometrie*, *Kennzahlen*, *Analyse* and the corresponding English terms *bibliometrics, indicators and analysis*.

The search engines used are Google Scholar and SpringerLink, which search for the above-mentioned search terms. In order to limit the text sources acquired in this way, criteria are selected that highlight potentially important articles. These are checked by means of backward and forward analysis for further publications to be examined. The following criteria were important in the selection process:

- Basics / Basic knowledge of bibliometrics
- Indicators
- Evaluation of indicators / criticism
- Current status of the publication
- Normalization

In order to understand and carry out bibliometric analysis, the basics of bibliometrics are necessary and a selection of indicators are required, which is why the publications are examined with regard to them. Furthermore, the identified indicators must be checked for suitability with regard to the object of investigation. Advantages and disadvantages must be weighed against each other. In order to be able to analyse across departments, one need information about the normalization of the calculated indicators (see Tunger, (2013), Ball, (2006)).

The following bibliometric indicators were used in our bibliometric analysis identified as a result of the literature search and finalized as a result of a discussion of the practicability of these indicators:

- Publications per Faculty
- Publications per Institute
- Publications by chair
- Publications per Author
- Citation rate
- Average citations per author
- h-index
- g-index
- hg index
- rational h-index
- Field normalized citation rate
- Percentage of publications cited by an author
- Percentage of uncited publications by an author
- Number of publications through collaboration.

Data collection

For the evaluation of an institution, a comprehensive set of indicators is needed to carry out the complex evaluation as accurately and objectively as possible. Each of these indicators requires a specific set of data that can be used to calculate the above-mentioned indicators. The analysis of the indicators showed that two distinct objects of investigation - the Author (A) and their publications (P) - are needed. The author includes first name, surname and the assigned subject areas, while the publication contains title, date, number of citations and the names of the authors. With the help of this data, the presented indicators can be determined almost completely. Table 1 shows an exact list of which indicators require which data.

Table 1: Data required by indicators

Indicator	*Necessary data*
Publications per Faculty	A, P,
Publications per Institute	A, P,
Publications by chair	A, P,
Publications per Author	A, P
Citation rate	A, P
Average citations per author	A, P
h-index	A, P
g-index	A, P
hg index	h-Index, g-Index
rational h-index	A, P
Field normalized citation rate	A, P, average citation per publication per field
Percentage of publications cited by an author	A, P
Percentage of uncited publications by an author	A, P
Number of publications through collaboration.	A, P

It must also be taken into account that departments or institutes of different faculties may have different communication habits that cannot be directly compared with each other and that cross-disciplinary analyses are highly negligent without considering the different publication habits. It is therefore indispensable to normalise key indicators in order not to distort the overall picture. The average citation habits required for normalization (field normalized citation rate) are determined using Web of Science. The search entry is differentiated according to the respective years and filtered according to the total number of German contributions. These search results are classified into the respective scientific categories such as chemistry or mathematics.

For the period 2012 - 2016, the search in all selected databases with the search terms for the address "Ilmenau" and with the wildcards "Il*me*au" to catch possible spelling errors delivered a total of 3138 publications. The data from the Web of Science was extracted on 04-02-2017 using a web crawler and stored as an XML file. Figure 1 shows an example of the XML publication structure.

```
<Publication Id="1" Title="Transitional boundary layers in low-Prandtl-number convection"
             Date="DEC 29 2016" Citation="0" ResearchArea="Physics"
             WoSCategory="Physics, Fluids &amp;amp; Plasmas">
  <authors>
    <Author firstname="Joerg" lastname="Schumacher" reference="1" />
    <Author firstname="Vinodh" lastname="Bandaru" reference="1" />
    <Author firstname="Ambrish" lastname="Pandey" reference="2" />
    <Author firstname="Janet D" lastname="Scheel" reference="3" />
  </authors>
  <addresses>
    <Address university="Tech Univ Ilmenau" referenceNumber="1" institute="Inst Thermo &amp;amp;
             Fluiddynam" />
    <Address university="Indian Inst Technol" referenceNumber="2" institute="Dept Phys" />
    <Address university="Occidental Coll" referenceNumber="3" institute="Dept Phys" />
  </addresses>
  <Info>By:Schumacher, J (Schumacher, Joerg)[      1 ] ;
             Bandaru, V (Bandaru, Vinodh)[        1 ] ;
             Pandey, A (Pandey, Ambrish)[         2 ] ;
             Scheel, JD (Scheel, Janet D.)[       3 ]
  </Info>
</Publication>
```

Figure 1: XML publication sample

Because the address details in the WoS are not stored in a standardised form and can differ from publication to publication, for example it is possible that one document contains the complete details of subject areas, institute and faculty, while others contain only the reference to the "Technische Universität Ilmenau", it was necessary to check each of the publications and to

precisely allocate them to the institutional units of TU Ilmenau. This mapping was achieved with a multi-stage procedure, the first source being the website of the departments of the individual institutes and faculties, in order to identify the current employees of the departments. In addition, existing electronic telephone directories for the years 2012 - 2016 were used to identify the employees for this period. As there were still a number of authors who could not be clearly assigned, an attempt was made to identify them with their institutional description using further scientific databases (SpringerLink...). All in all, 2907 publications (99.35%) with a specific assignment of author, publication and institution could be made available for analysis.

Data processing and analysis

The data for bibliometric analysis is mapped and stored in a relational database system to ensure sustainability. The powerful query and report functions simplify the evaluation of the data. The Access 2016 database consists of the two main elements *author* and *publication*. The *university* table refers to institutional elements such as faculties, institutes and departments. The semantics of the manifold m:n relationships (e.g. such that a department belongs to several institutes, while an institute can consist of several departments, staff-department, staff-publication) leads to a series of further intermediate tables, so that the final database consists of a total of 21 tables, twelve tables containing all the necessary data to store the publications and the authors. The other tables serve to store the calculated indicators.

Results

This section presents and compares distinct results of the respective indicators for the bibliometric analysis of the publication behaviour of the TU Ilmenau.
The publications are analysed on four levels of aggregation. First, the faculties as a whole are examined and the associated publications are summed up. The same is done with the institutes, the departments and the authors. Figure 2 illustrates the overall publishing rate for each faculty

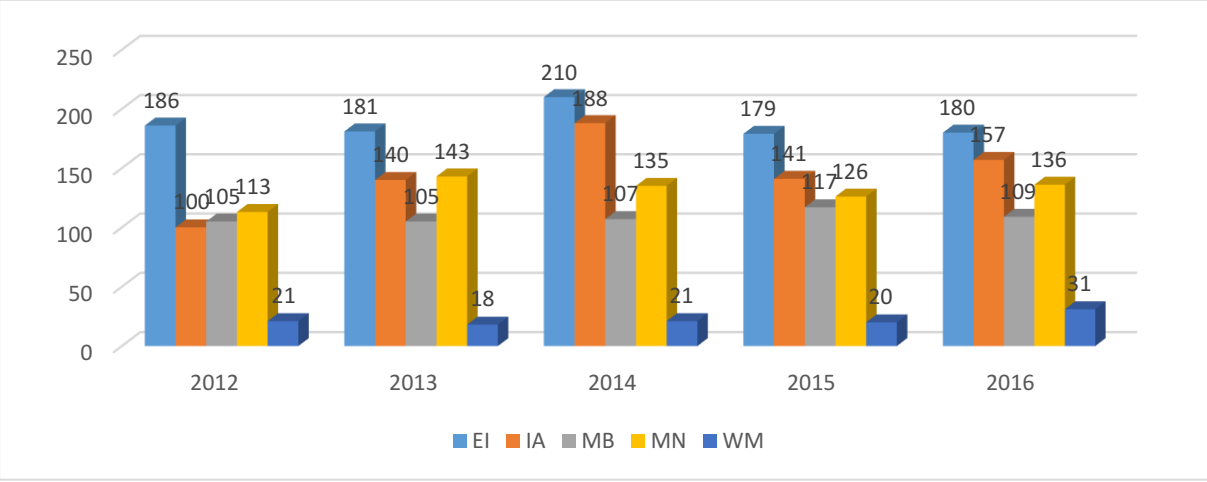

Figure 2: Publishing rate per faculty

Table 2 illustrates the top ten institute according to the publishing rate. As can be seen and as was to be expected for a Technical University, the publication landscape is dominated by the technical and scientific institutes.

Table 2: Top ten institutes (publishing rate)

	Institute	*total*
1	Institute for Information Technology	460
2	Institute for Physics	353
3	Institute for Computer and Systems Engineering	221
4	Institute for Microelectronics and Nanoelectronics	173
5	Institute for Chemistry and Biotechnology	170
6	Institute for Biomedical Engineering and Informatics	161
7	(Inter-departmental) Institute of Materials Science and Engineering	142
8	Institute for Thermodynamics and Fluid Mechanics	140
9	Institute for Theoretical Computer Science	135
10	Institute for Mathematics	130

The Technical University of Ilmenau lists 138 departments assigned to the respective institutes. There are five departments, which have published more than 100 publications during the whole period of the study.

Table 3: Top departments with more than 100 publications

	Departement	*total*
1	Electronic Measurement Engineering Group	154
2	Biomedical Engineering Group	146
3	Communications Research Laboratory	120
4	Group for Complexity Theory and Efficient Algorithms	116
5	Chemistry Group	107

A total of 2907 publications were published by 1314 authors. Table 4 lists the top five authors with the highest number of publications.

Table 4: Top five scientists (publishing rate)

	Scientist	*Total*
1	M D	110
2	M H	105
3	J H	100
4	R T	97
5	U R	80

Table 5 shows a list of the top 5 cited papers from different academic fields.

Table 5: Top five cited papers

	Paper	*Total citation*
1	Controllable Disorder Engineering in Oxygen-Incorporated MoS2 Ultrathin Nanosheets for Efficient Hydrogen Evolution	427
2	Vacancy Associates Promoting Solar-Driven Photocatalytic Activity of Ultrathin Bismuth Oxychloride Nanosheets	255
3	Binary copper oxide semiconductors: From materials towards devices	144
4	Graphene Transistors: Status, Prospects, and Problems	136
5	First-principles investigation of the size-dependent structural stability and electronic properties of O-vacancies at the ZnO polar and non-polar surfaces	120

In order to make the results from different scientific fields comparable with each other, the absolute values of the publications were normalized with the field-normalized citation rates, in keeping with the field-dependent publication habits and thus enabling an objective comparison. A result greater than 1 therefore means an above-average value for citation per publication in Germany. The citation habit for the Institute of Physics for 2013 is 13.68. The publication of the *Three-Dimensional Nanostructuring Group* of the *Institute of Physics*:"*Controllable Disorder Engineering in Oxygen-Incorporated MoS2 Ultrathin Nanosheets for Efficient Hydrogen Evolution*" thus received more than 31 times the average citation. Using this standardization approach, it was possible to compare indicators based on the publication output like h-index, g-index or hg-index of scientists throughout the university of different research areas like *Chemistry*, *Mathematics* or *Economics*.

At this point, we do not present in detail the results of the other publication-related indicators in preference to two indicators that can also be interpreted as qualitative. The ratio of cited publications to the total number of publications can be seen as a quality criterion for scientific output if we assume a certain basic quantity of publications. Of 1314 authors, 882 have at least one citation for a publication they have produced. A total of 431 scientists who have published between one and 13 publications are not cited. 219 authors have published a publication that has also been cited. 87 authors have published a publication that has been cited, as well as a publication that has not been cited. In the case of 85 scientists, exactly one cited publication is compared with 2 to 52 publications that were not cited. It is interesting to compare authors who have published more than 50 publications. The citations are put in relation to the total number of publications. Figure 19 shows this ratio for authors with a number of publications > 50.

Table 6: Scientists with the highest ratio of cited papers with more than 50 publications

	Scientist	*cited*	*uncited*	*ratio (%)*
1	H H	51	9	85
2	Y L	62	11	85
3	M K	46	10	82
4	P S	52	12	81
5	U R	61	19	76
6	A B	38	15	72

Further interesting conclusions about the way of scientific work can be made by analysing the collaboration behaviour of authors. For this purpose, the number of publications that were created in co- or multiple authorship was measured. The number of publications created through collaboration is intended to show how often employees of the TU Ilmenau cooperate. No difference is made here between cross-university and intra-university publications.

Table 7: Top five scientists in multiauthorship

	Scientist	*Total*
1	M H	105
2	R T	97
3	U R	80
4	H H	60
5	F R	57

We have further investigated these collaboration networks in further examinations. Figure 3 shows the result of a visualization of the collaboration network with VOSViewer[1].

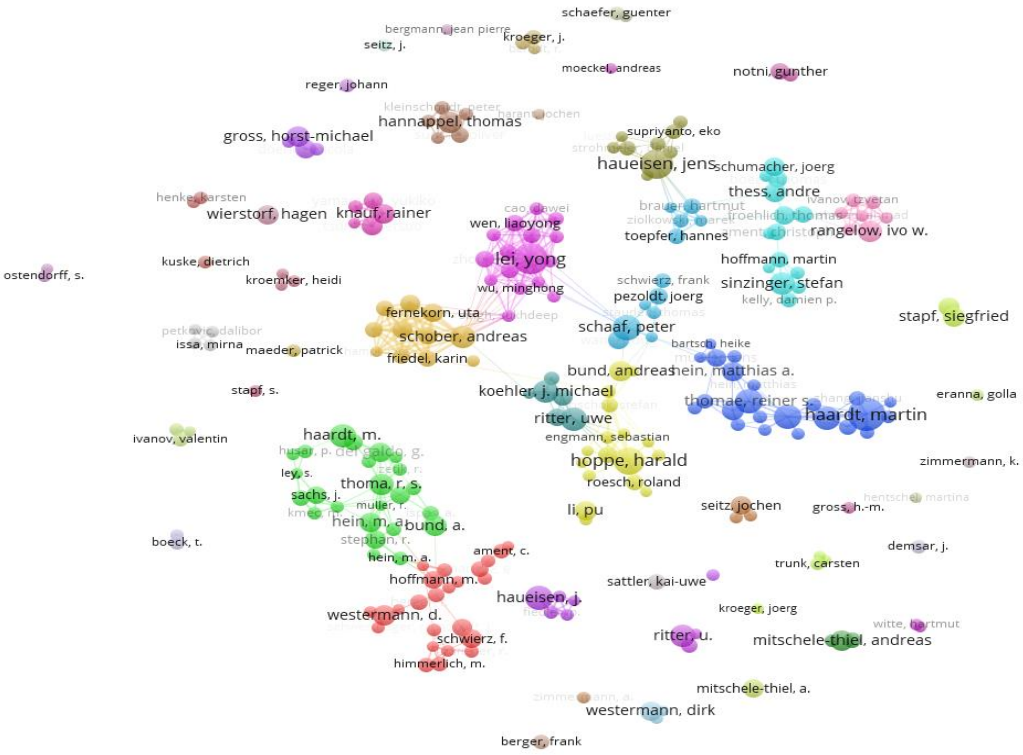

Figure 3: Co-Authorship Network of TU Ilmenau 2012 – 2016 (threshold 10)

Obstacles

The correctness of the underlying data material is a requirement for high-quality bibliometric analysis. In our specific example we had to deal with a number of error sources typical for such bibliometric investigations.

Accuracy of the data

- The data required for the bibliometric analysis were not always available in correct form. The author names were neither available on the website of the Technical University of Ilmenau, nor on the Web of Science, where they were one hundred percent error-free. Since each department maintains its own employee list online, there was no standardized form of representation and an automated extraction of the names was therefore not possible (use of different separators in the name separation). Especially with foreign names, there were often uncertainties as to which parts belong to the first name and which to the last name. As long as not all authors are registered in standardized repositories (ORCID, Researcher-ID or Scopus Author ID or similar), the solution can only consist of time-consuming manual post-processing. In our case, the author data filtered out of the website were compared with those of the telephone lists. If a person from the telephone lists did not exist with a complete first and last name in the employee list, the system searched for the last name and the first letter of the first name. If a hit was found, we manually checked whether this corresponded to the correct person. Similar problems with the assignment of names occurred in the Web of Science.

[1] The detailed outline of the results of the data science analysis of the available data sets will be the content of another publication in the pipeline.

Incorrectly written names were checked manually. For this purpose, a list was programmed to save these exceptions and then transferred it to the Access database.

- An essential requirement for the correct automated extraction of data from the Web of Science, was the correct pre-structuring with appropriate separators (bsw. comma between first and last name). If such a separator does not exist, the corresponding attribute in the raw XML file remains empty and the assignment must be completed manually. Especially often such a manual check was necessary for names from other cultures.
- For a number of publications, the authors contained references to institutional units that could not be assigned to them. In order to be able to guarantee the affiliation, all Ilmenau-related authors of a publication were first collected. The authors were then compared with the existing staff database (consisting of the data from the TU Ilmenau website and the telephone lists for the years 2012 to 2016). If there was an entry with a complete first and last name, the search was based on the number of departments. If this search yielded only one result, it was assigned to the employee. Otherwise, this data was collected and later assigned manually. A tool was programmed to select the right authors and subject areas.

Duplicate entries

- When collecting the data, it was found that publications (either as e-books or as articles available elsewhere) were published twice in the Web of Science. In order to filter these out, identical publications were searched for and the publication date as well as the names of the authors were checked. If they matched, all identical publications were combined into one. In concrete terms, this means that one publication was removed from the database and the citations were grouped.

Missing data

- Some addresses were not available in the publications, and therefore had to be added, i.e. also that no references were available. In order to correct this, the employee list was searched for hits and the author names were additionally checked by an additional Google search to ensure assignment. The author names were examined with the help of third-party sources (such as SpringerLink and Google Scholar) in which the reference to the corresponding addresses was available and an assignment of the institutional affiliation ("affiliation") could be carried out.

Summary and Future Work

With the help of a bibliometric analysis, a cross-disciplinary comparison of the publication performance of the TU Ilmenau was carried out. The available results allow the publication output of the past five years to be analysed and the strengths and weaknesses of scientists and their corresponding organisational units to be uncovered. There are considerable differences in publication behaviour between the different university tiers. A number of factors that have contributed to these results are relevant for the present results. The size of the departments and institutes can be an advantage over the smaller sized units and can have a direct influence on the position in the respective ranking, since the number of actively publishing employees can correlate with the amount of publication output within a period. By means of normalization, the indicators of specific groups were made comparable despite different citation habits. In this paper, a series of indicators were used to describe the research landscape at TU Ilmenau as objectively as possible. The Data acquisition was one of the most complex parts of the work. A number of programs were developed to convert the raw data into a usable form suitable for

bibliometric analysis. In order to provide consistent and correct data for the analysis, the data was manually checked in several stages (after each processing step) in order to minimize potential sources of error. For the future, further use of the data, they were stored in a structured form in a database.
This database can be used as a basis for further investigations. It can serve as a basis to re-examine publication behaviour at regular intervals, to draw comparisons and to work out trends and tendencies. It was also used for the work currently being completed on more content-oriented analyses, which have a more Data Science-specific character, to answer such research questions like: If there are changes in research areas, are there overlapping research areas that offer the potential for new opportunities for scientific cooperation, or how detailed the research landscape of a university can be analysed by including more content specific elements like keywords and abstracts.

References

Alonso, S.; Cabrerizo, F.J.; Herrera-Viedma, E.; Herrera, F. (2009) h-index: A Review Focused in its Variants, Computation and Standardization for Different Scientific Fields. *Journal of Informetrics*, 3(2009)4, pp. 273-289.

Ball, R. (2006). Der Hirschfaktor - ein neuer “Stern” am bibliometrischen Indikatorenhimmel. *B.I.T.-Online*, 9(2006)4, pp. 309-312.

Ball, R. (2007). Wissenschaftsindikatoren im Zeitalter digitaler Wissenschaft. *B.I.T.-Online*, 10(2007)2, pp. 121-125.

CWTS *Leiden Ranking* (2017). Retrieved December 19, 2017 from: http://www.leidenranking.com/ranking/2017/list.

Havemann, F. (2009). *Einführung in die Bibliometrie*, Berlin, 2009. 66 p. Retrieved February 7, 2018 from http://www.wissenschaftsforschung.de/Havemann2009Bibliometrie.pdf,

Schulze, N.; Michels, C.; Frietsch, R.; Schmoch, U.; Conchi, N. (2013). *Indikatorbericht Bibliometrische Indikatoren für den PFI Monitoring Bericht.* 2013. 85 p. Retrieved March 7, 2018 from: https://www.isi.fraunhofer.de/content/dam/isi/dokumente/ccp/2012/Indikatorbericht_PFI_2012.pdf.

Shanghai (2017). *Academic Ranking of World Universities* 2017. Retrieved December 19, 2017 from: http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/.

Tunger, D.: *Möglichkeiten und Grenzen bibliometrischer Analysen*. Forschungszentrum Jülich, 2013

Webster, J, Watson, R.T.: Analyzing the Past to Prepare for the Future:Writing a Literature Review. *MIS Quarterly*, 2(2002)26, pp. xiii-xxiii.