

Analysis of Publishing Activities and “Publish or Perish” Strategy

VACLAV SKALA

Department of Computer Science and Engineering

University of West Bohemia

Univerzitni 8, CZ 30614 Plzen

CZECH REPUBLIC

<http://www.VaclavSkala.eu>

Abstract: - An “objective” evaluation of research results is a very difficult task and many very sophisticated approaches have been reported. The current trend tends to “record counting” and summation of “points” gained. It leads to a long term deformation of a researcher’s activities especially if she/he is evaluated on this principle and her/his salary depends somehow on the number of “points” gained.

In the Czech Republic such system of “counting” was introduced recently and university budget depends on such “points counting” significantly. Actually this approach has been applied in the Czech Republic during the last 10 years.

This paper presents some indications of possible development in future as a similar approach can be seen through all over the world last years under “Publish or Perish” term. However, this approach seems to be a disaster in a longer time span as a pressure to publish at an individual level leads to concentration on a short term research activities.

Key-Words: - Publication activities analysis; research evaluation; trends in publications; business model

1 Introduction

There was a great invention in history that changed our civilization – a paper making process in China, which can be dated back to 2nd century BC in China. The second great invention was made by Johannes Gutenberg who started the “Printing Revolution” in the 15th century. The both inventions caused incredible development of our civilization not only from the cultural and technical point of view, but generally. However the computer principle invented by Charles Babbage in 1822 was a principal invention which waited for real applications and its use till the 20th century. In the late 1960s there was acceleration by the ARPANET/DARPA research activities resulting into Internet development.

Nowadays, the current technology enables not only a flexible download of publications and their printing (also a 3D print is affordable at home, now), but also to produce publications very fast and much more cheaply than 20 years ago.

In research and science oriented activities, a researcher or a research team is evaluated mostly on previously reached results, which is actually a short term evaluation. Grants are allotted according to contents of the project proposal and recent research and publications made. A final research grant evaluation is mostly based on

publications actually made at the time, when the grant or project ends. As grants or projects are usually of the 3-4 years span, and the period from a paper submission to its publication is usually about 1 year, researchers are under a significant time stress. They are indirectly forced to submit their results to related conferences. This factor is even more severe in the case of PhD students, of course.

Those factors have been contributing to an explosion of conferences worldwide significantly. There are also other factors that stimulate conference organization as it makes the universities involved more visible on a public space.

An objective evaluation of research results is a difficult task and many very sophisticated approaches have been reported, including approaches based on scientometry etc.

However the current trend is “record counting” and summarizing “points” gained, which leads to a long term deformation of a researcher’s activities especially if she/he is evaluated on this principle and her/his salary depends somehow on this.

In the Czech Republic such system of “counting” was introduced recently and university budget depends on such “points counting” significantly. This approach has been applied in the Czech Republic during the last 10 years.

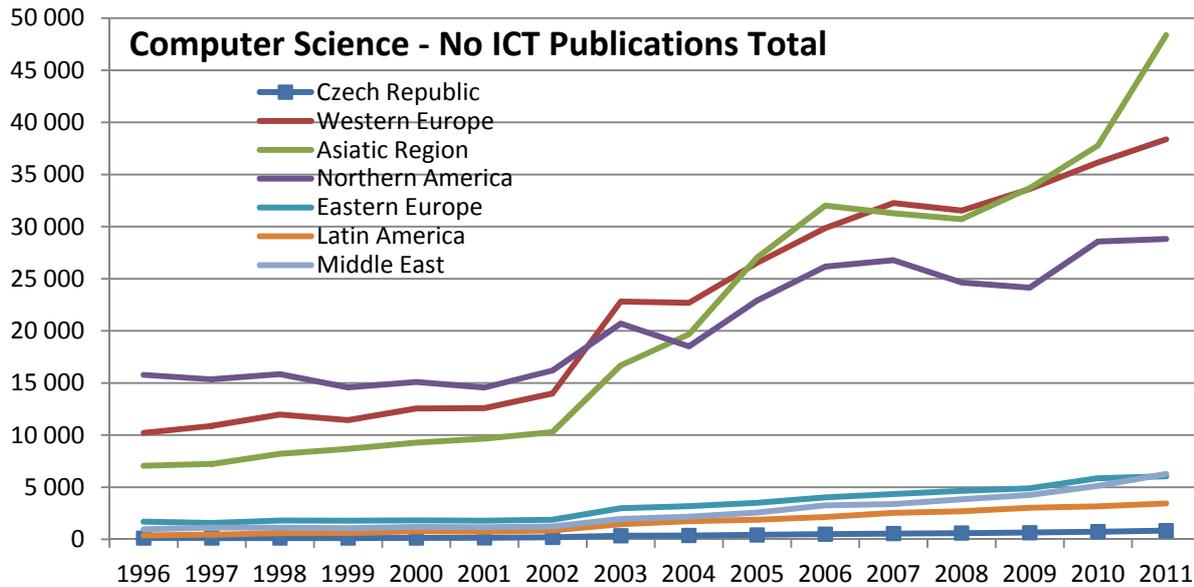


Fig.1 Number of publications worldwide – SCOPUS / Scimago

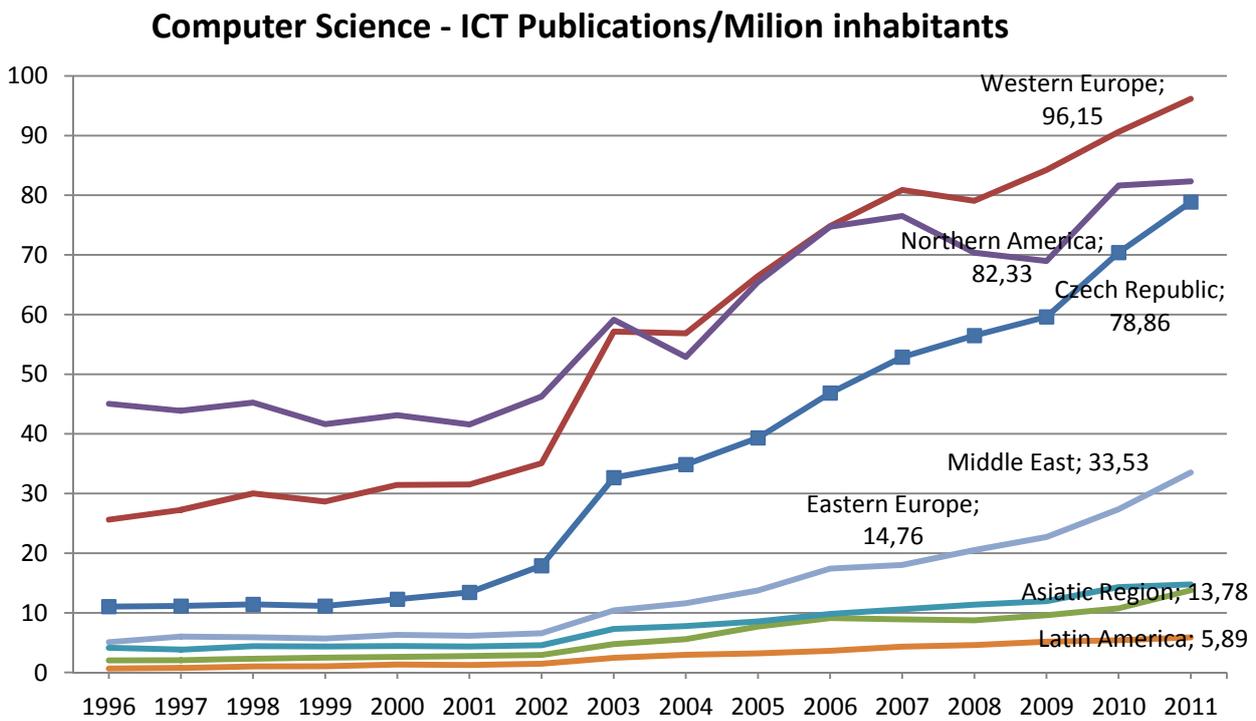


Fig.2 Number of publications worldwide per one million of inhabitants – SCOPUS / Scimago

This paper presents some indications of possible development in future as a similar approach can be seen through all over the world last years under a term “Publish or Perish”.

However, it seems to that this approach can lead to a disaster in fundamental research in a long term extent. It can be seen that there is an explosion of papers published nowadays and even in a very narrow research field there is no chance to read or

overview all the relevant papers published. Another quite interesting feature of today’s publications is that there are, in some cases, “closed small research communities” having common journal, conferences or workshops and citing themselves mutually to increase their citation indices etc.

In this paper we will show some interesting results of a simple analysis of publication data publicly available.

2 What Data Can Say

There are a lot of studies on publication activities, their influences and comparative studies. Usually those statistics present how many papers, citations or patents have been made per country etc. This is of course a totally wrong approach as all indicators have to be relative, e.g. to a number of inhabitants or to a number of university employees etc.

In the following we selected Computer Science field for experiments. Data from different sources were used, e.g. SCOPUS, ISI-Thomson Reuters and Scimago etc. The SCOPUS and Scimago databases were used, unless explicitly said otherwise.

Fig.1 presents total number of publications in different regions and in the Czech Republic. It can be seen that regions Western Europe, Northern America and Asiatic Regions are highly documents productive, while other regions not. However for a real evaluation the relative criteria are to be used.

Fig.2 presents a relative chart, i.e. Number of publications per one million country/region inhabitants. This graph clearly shows trends in different regions and in the Czech Republic. According to data available the Czech Republic is about on 80% of the Western Europe publication level.

There is clear positive influence of the publication evaluation system introduced 10 years ago. It means that on the “macro” level it can be considered as a positive strategy, however on the “individual” level, i.e. on the individual evaluation, this strategy is clearly contra productive.

However there are differences across different fields within computer graphics. Tab.1 presents publishing activities in different fields within Computer Science area.

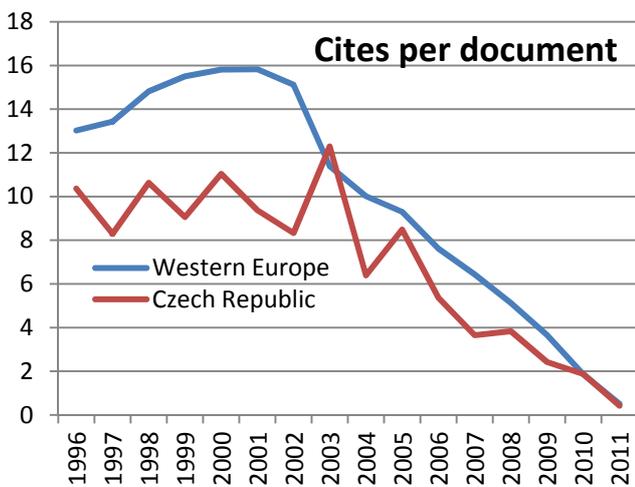


Fig.3 Cites per a document ratio

There is another significant threat coming with the “Publish or Perish” strategy. As number of

publications grows non-linearly every year, e.g. for Information systems grows with $O(N^2)$ complexity in general, researchers usually have “constant” number of references they used during their research. Therefore number of citations per all documents decreases with a ratio $O(1/N^2)$. Analysis of available data supports this hypothesis, Fig.3.

Another factor is the case of self-citing. There are some objective reasons for that as researchers should somehow refer to their previous publications where additional detailed or relevant information can be found. However, there are significant different attitudes to self-citations in general, see Fig.4 (only countries with more than 500 papers in computer science in 2006-2011 period were considered).

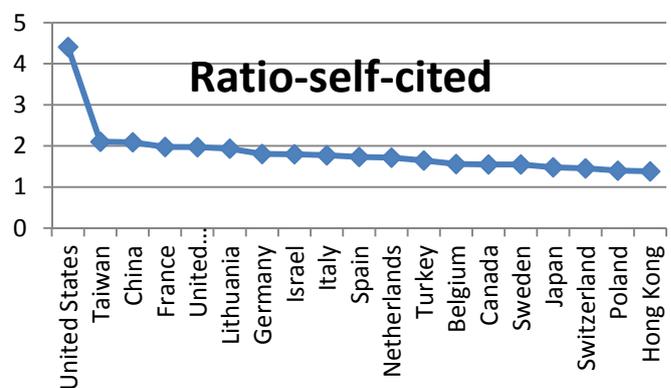


Fig.4

The graph at Fig.4 should be read as follows – if a paper has 10 references listed, then 1,2 – 4 references are the authors self-cited references.

Another interesting aspect is the ratio of ICT publications and all publications per a country (only countries with more than 500 papers in the time span 2006-2011 were considered), see Fig.5. It gave us clearly information how ICT fields are important against other fields. Of course, small countries have to be taken with some “tolerance” due to small samples used. However, countries like Hong Kong, Singapore, Taiwan are leading in ICT industry. On the other hand Japan and Czech Republic are very similar as far as the ICT/ALL publications ratio is concerned.

Let us consider the ICPT publications only, now and let us explore how different fields within computer science are “productive” in publications. Tab.1 presents publishing activities to 1 million of inhabitants of a country or region relatively. The profiles of publications in the Czech Republic closely follow Western European countries due to

Ratio ICT/ ALL

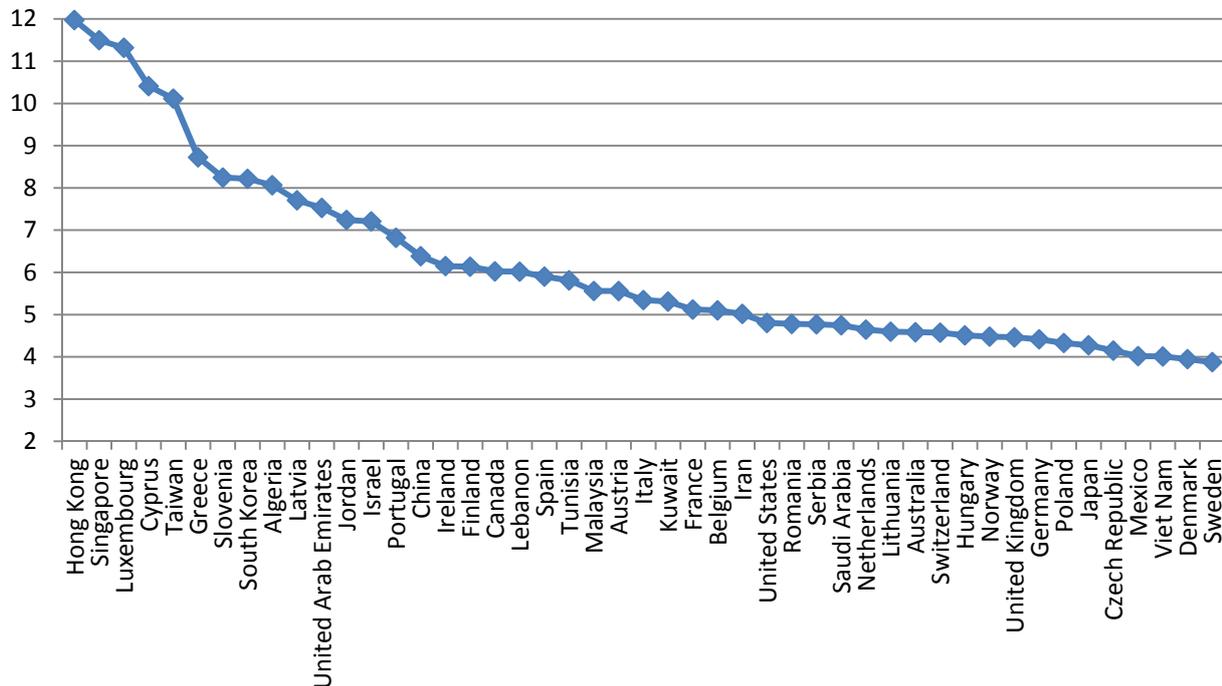


Fig.5 Ratio of ICT publications/ all publications

very strong economical interlink between Czech Republic and European countries.

This gives us basic information more or less on research activities as follows:

- **Computational Theory and Mathematics (CT&M)**, Tab.1.a: USA & Canada, Western Europe and Czech Rep. follow the same steadily growing trend. This is probably given by challenges in civil, space and military industries.
- **Computer Networks and Communications (CN&C)**, Tab.1.b: USA & Canada and Western Europe have the same trends; however research activities in the Czech Rep. are much lower probably due to lower level of research activities in microelectronics.
- **Human Computer Interaction (HCI)**, Tab.1.d: activities in this field are significantly higher than in Western Europe or USA & Canada. However detailed analysis shows that it is not the case as far as papers with impact factors.

2 Impact of the Selection Process

So far we used SCOPUS and Scimago databases. However in the Czech Republic evaluation is ISI/Thomson Reuters database preferred. Fig.6 presents number of publications from physical sciences field.

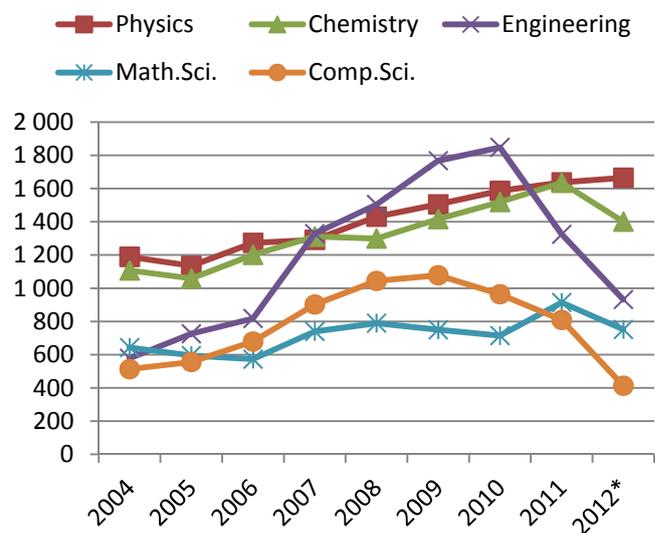
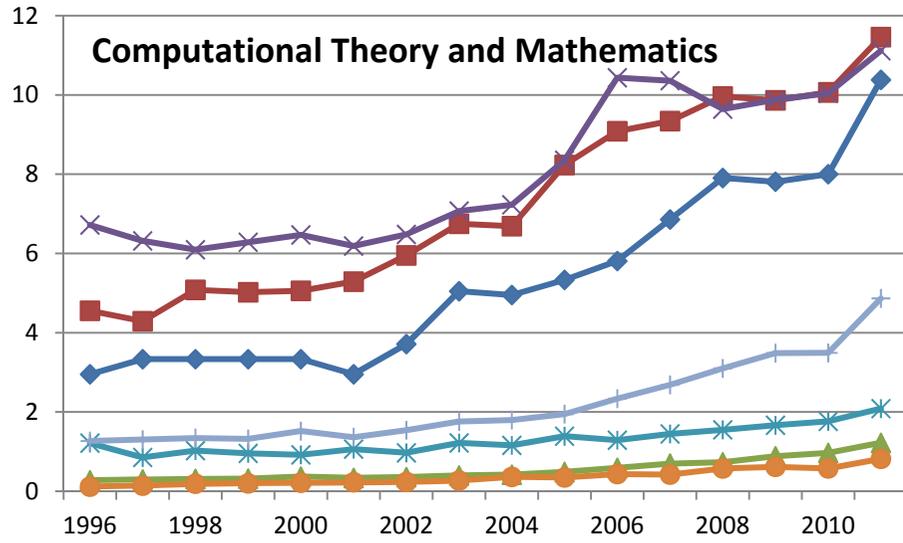
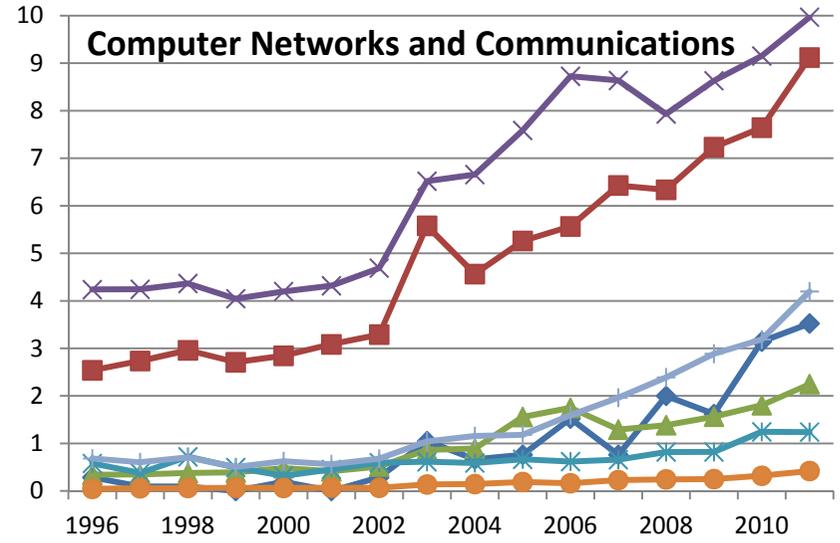


Fig.6 Number of publications in physical sciences in ISI/Thomson Reuters (retrieved on 2013-01-13)

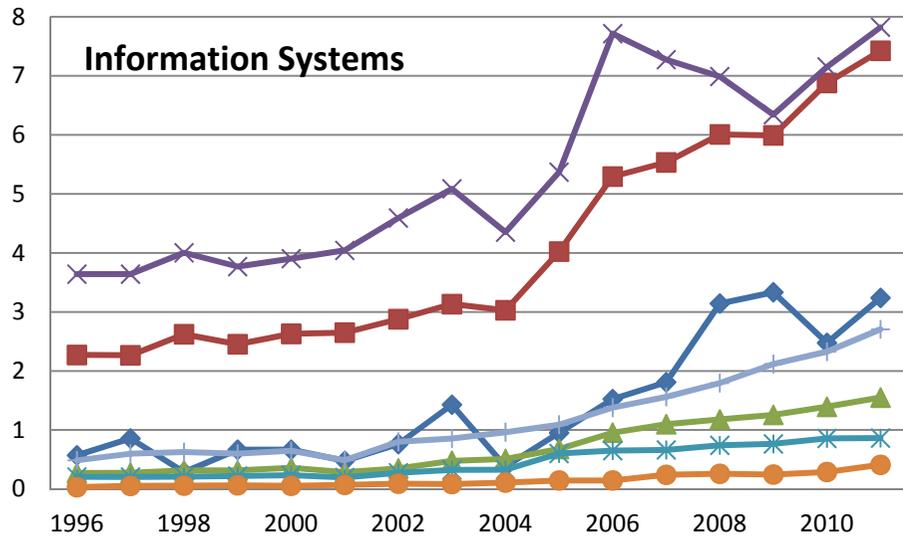
The graph is actually saying that number of publications in computer science and engineering is decreasing according to ISI/Thomson Reuters. It does not correlate with the SCOPUS database. There is a hypothesis that there was a change in a selection policy in 2009 which is probably stricter, now. However selection criteria remain unknown.



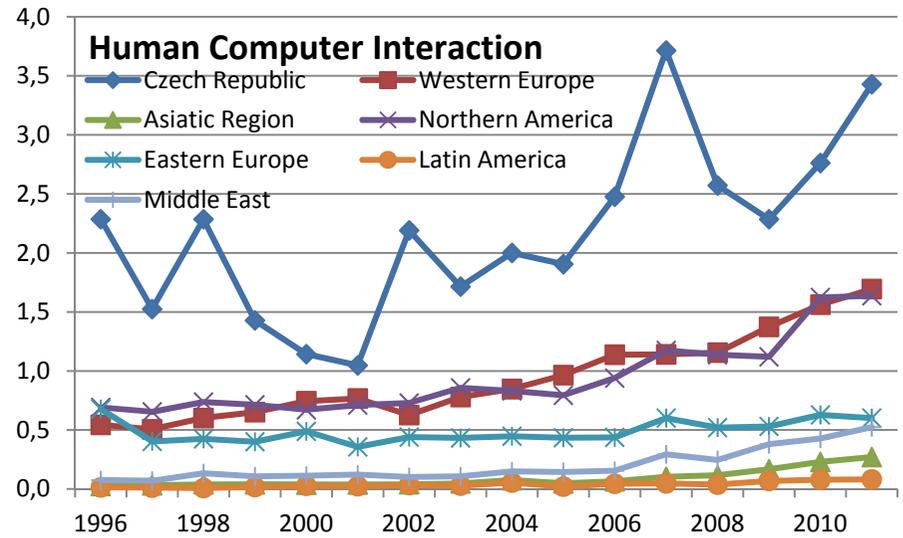
a)



b)



c)



d)

Table 1 Publication activities in Computer Science

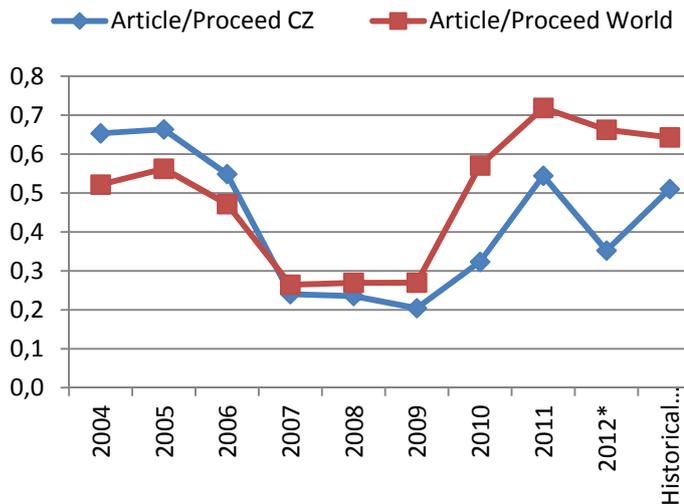


Fig.7 Ratio of “Article”/“Proceedings paper” in ISI/Thomson Reuters (retrieved on 2013-01-13)

Fig.7 presents the ratio of “Articles”/“Proceedings papers” retrieved from the ISI/Thomson Reuters database for physical sciences. It can be seen that in the period 2007-2009 there was a significant increase of proceedings papers indexed by ISI/Thomson Reuters. This supports the hypothesis that stricter policy in indexing proceedings papers was introduced in 2010. However, unless the selection criteria are publicly known and they should be based on paper’s scientific contribution measures, the ISI/Thomson Reuters should not be consider as major resource for evaluation..

3 Current Situation

Let us summarize the most significant factors of today’s situation:

- Researchers and PhD students are under constant stress to publish research results in journals with an impact factor or in conference proceedings.
- There is a limited time span for actual research work and preparing a publication and getting it published.
For non-institutionally funded grants, the period granted for funding is usually limited to 3-4 years.
- Researchers and PhDs are evaluated according to a number of publications with some evaluation of quality, usually based on the impact factor or if the publication is indexed by some commercial database, like Scopus, ISI-WoS etc.
- There is a need to distribute financial means for research and development in a “meaningful”

way. It means that some “measure” of a contribution of a publication is needed.

However research and publication evaluation as it can be seen nowadays definitely leads to “short money”, i.e. financial means are allocated for a short period, while fundamental research, for which such financial means are to be actually targeted, need “long money”, i.e. funding for a long period.

There is another very significant factor which can be detected during the last decade – there are so many declarations of research and innovations needs, but the budgets allocated to fundamental research are more or less constant or it decreases. It seems that politicians use this as “buzzwords” for their personal public promotion without responsibility to research communities and to fundamental research itself.

3 Time Factor

There is one surprising factor found in data analysis made. Fig.8 presents absolute number of publications in different fields, while Fig.9 presents relative growths from the year 1996 to 2011. It can be seen that number of publications in “Medicine” is absolutely very high, while “Computer Science” is the most dynamically growing except of “Arts and Humanities”

4 Publication Business

As there is a “fiction of need” to get something published in order to make a promotion, PhD etc. there is a new opportunity for business activities in publishing field.

In the past, there was a lengthy and costly procedure as the manuscript had to be rewritten, reset, corrected and finally when got published, the paper had to be sent by mail to a reader.

Nowadays the author is expected to prepare manuscript, correct it according to reviews, if any, and make the final typesetting using TEX or MSWord etc., convert it to PDF or similar format. A publisher includes the paper into the database and makes it available. Also couple of printed issues at a “horrible” cost is distributed by air-mail. If a paper is downloaded from the database, usually an individual cost is about 10-30 EUR per paper.

A new business model is based on “open access” principle, which actually means that the author or her/his institution has to pay approx. 500-1000 EUR per paper, if accepted for a publication, to get it published.

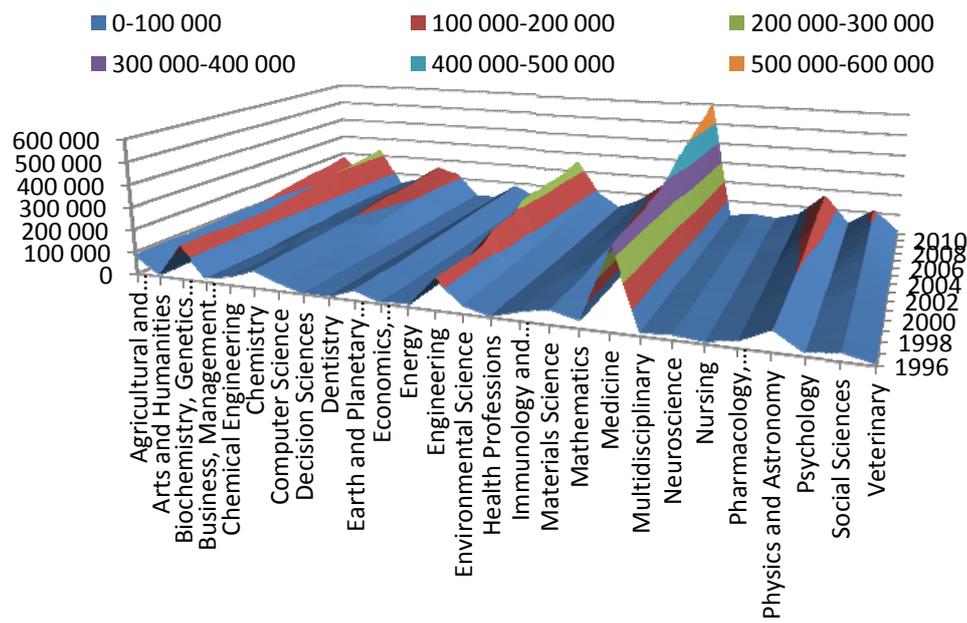


Fig.8 Absolute number of publications in 1996-2011 period

Ratio 2011/2006

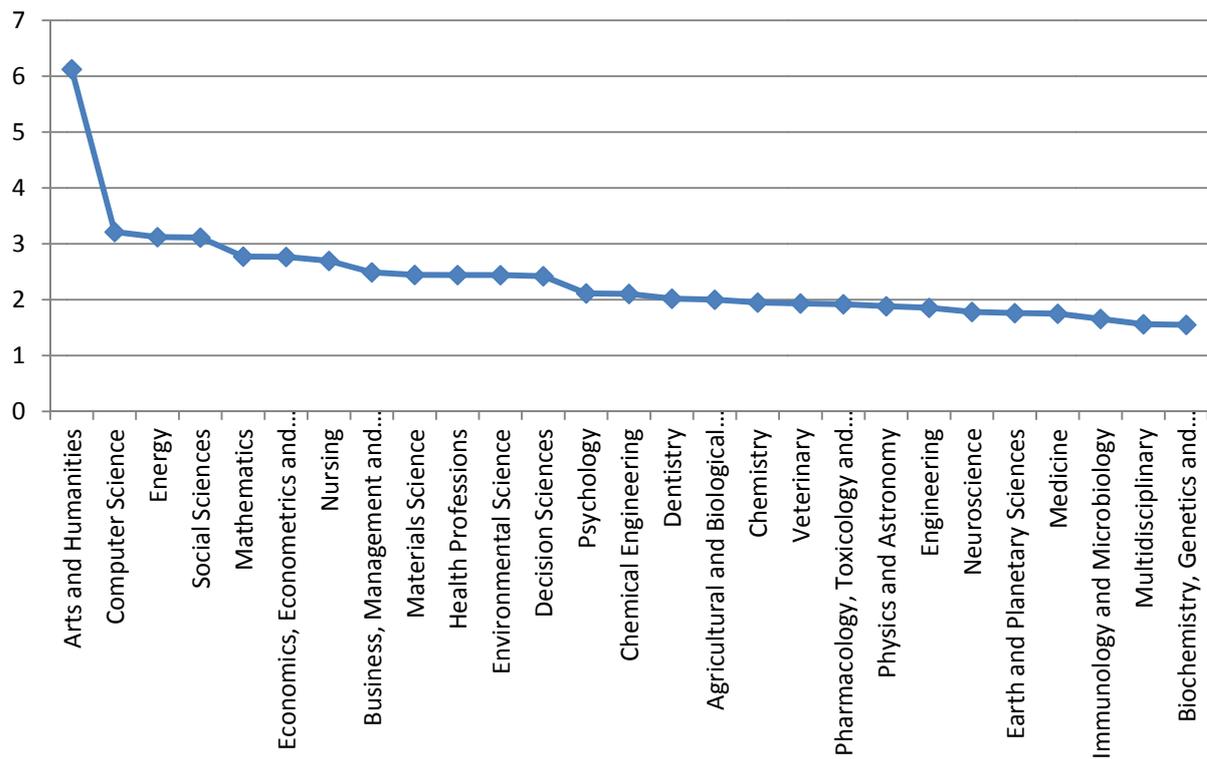


Fig.9 Ratio of publications 2011/ 2006 period

It actually means that the business model is similar to conference organization and proceedings production more or less, now.

From the “new publishing” company point of view: to be successful in the publishing business requires meeting the following requirements:

- find out “hot topics” for the publishing activities
- attract perspective authors world wide
- decrease “submission – review - get published” cycle as much as possible
- get it indexed by indexing companies or by global electronic indexing systems like scholar.google.com, arnetminer.org or academic.research.microsoft.com etc.

This is actually a trend that can be seen nowadays. There is an explosion with new conferences, workshops etc. promising to get the paper or its extended version published later on.

From the authors point of view: there is a legitimate question whether “reasonable” publications cannot be published by authors directly if they find how to disseminate publications somehow and get them included into world indexing systems.

It seems to that there is a new era of information dissemination approaching with systems like Digital Public Library [7] or Google Books etc. This might be somehow reducing the cost of research results publishing and cost of retrieving papers form digital repositories.

5 Conclusion

This paper presents a summary overview of selected aspects of publications indexed by Scopus, Scimago, Thomson Reuters etc. The reason of this paper is to show how some evaluation aspects acceptable at a macro level might be danger if used on an individual level and how if used improperly in allocation of financial means for research can be harmful for long term fundamental research for which the funding actually aiming to.

Acknowledgment

The author would like to thank to colleagues at the University of West Bohemia, Plzen and VSB-Technical University, Ostrava for their comments, suggestions and hints that contributed to this study. Thanks belong also to anonymous reviewers for their comments and for some inconsistencies pointing in their reviews.

The activity was supported by the Ministry of Education of the Czech Republic, projects No.ME10060 and LG13047.

References:

- [1] SCImago. SJR — SCImago Journal & Country Rank. <http://www.scimagojr.com>
Retrieved <2013-02-04>
- [2] SCOPUS – <http://www.scopus.com>
Retrieved <2012-02-04>
- [3] Thomson Reuters – Web of Knowledge
<http://www.webofknowledge.com>
Retrieved <2012-02-04>
- [4] <http://academic.research.microsoft.com>
Retrieved <2012-02-04>
- [5] <http://www.arnetminer.org/>
Retrieved <2012-02-04>
- [6] <http://scholar.google.com/>
Retrieved <2012-02-04>
- [7] Thompson,Ch.: Read Free of Die, IEEE Spectrum, pp.50-51, January 2013