Impact Factor:			
ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	РИНЦ (Russia) = 0.234	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 1.042	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

SOI: 1.1/TAS DOI: 10.15863/TAS

International Scientific Journal

Theoretical & Applied Science

p-ISSN: 2308-4944 (print) **e-ISSN:** 2409-0085 (online)

Year: 2017 **Issue:** 02 **Volume:** 46

Published: 28.02.2017 **http://T-Science.org**

Petr Alekseevich Boldyrev
Director of Research Library
Orenburg State University,
Russia, Orenburg
library@mail.osu.ru

SECTION 4. Computer science, computer engineering and automation.

THE APPROACH TO THE CALCULATION OF SCIENTOMETRICAL INDEXES ON THE BASE OF DATA FROM RUSSIAN AND FOREIGN QUOTATION SYSTEMS UNDER THE CONDITIONS OF NONCOMPLETE DETERMINATION

Abstract: *This work presented the problem of calculation of the main scientometrical indexes on the base of data from different quotation systems under the conditions of noncomplete determination caused by limited access to the quotation systems. It was offered the approach including following components: data collection and processing; the construction of aggregative publications list; the calculation of total number of quotations for each publication; the calculation of citation index and Hirsch index. The offered approach allowed make the calculation of citation and Hirsch index on the base of data from RSCI and SCOPUS quotation systems under the conditions of noncoplete determination. There were carried out the investigational studies of developed approach that allowed make the conclusion about the possibility of their usage for calculation of the citation and Hirsch indexes under the conditions of noncomplete determination.*

Key words: *scientometrical indexes, citation index, Hirsch index, conditions of noncomplete determination, the construction of aggregative publications list, the calculation of total number of quotations.*

Language: *English*

Citation: Boldyrev PA (2017) THE APPROACH TO THE CALCULATION OF SCIENTOMETRICAL INDEXES ON THE BASE OF DATA FROM RUSSIAN AND FOREIGN QUOTATION SYSTEMS UNDER THE CONDITIONS OF NONCOMPLETE DETERMINATION. ISJ Theoretical & Applied Science, 02 (46): 205-208.

Soi: http://s-o-i.org/1.1/TAS-02-46-35 ***Doi***: crossref https://dx.doi.org/10.15863/TAS.2017.02.46.35

Introduction

At the moment to estimate the effectiveness of scientific work the scientometrical indexes are used together with experts' opinions. It's due to the presence of secure available for measurement and comparison information about the scientific researches results presented in different quotation systems. To the most popular foreign quotation systems can be referred Web of Science [1] and Scopus [2] and to the Russian – RSCI [3].

However despite the big choice of quotation systems offering data for estimation of authors' publication activity there is number of problems preventing their wide usage in scientific and educational organizations [4]. Quotation systems Web of Science and Scopus don't include the majority of publications in Russian. Quotation system RSCI doesn't have an access to the big number of foreign publications and also doesn't have the majority of works up to 2000. Special importance this problem gets under the conditions of limited access to scientometrical information because of high price.

Materials and Methods

Some scientists tried to develop software tools providing additional possibilities working with quotation systems [5, 6, 7]. However after the analyses [8, 9] of developed software tools following conclusions were made:

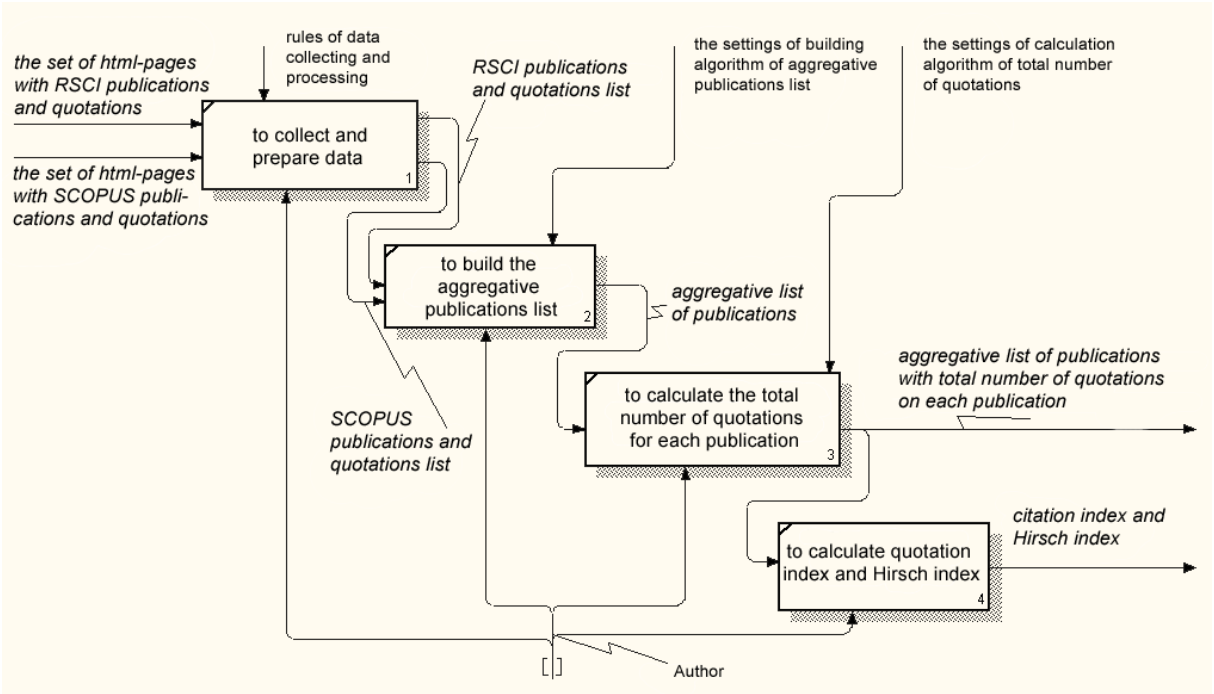
- there aren't software tools that can aggregate data and calculate scientometrical indexes taking into account Russian Science Citation Index;
- developed software tools are supposed to work with the availability of full access to the quotation systems and can't work under the conditions of noncomplete determination.

In that case under the conditions of noncomplete determination is meant the absence of possibility to identify uniquely and relate the

quotations (the title of quoted work, the source and publication year) in foreign quotation systems with the list of quotations in RSCI for each publication. Whereas there is provided only the quantity of quotation for each chosen publication in foreign quotation systems. As a consequence there is the problem of calculation of total quotations number for each publication. It's also impossible to calculate the main scientometrical indexes (quotation index, Hirsch index [10] and others).

Consequently the objective of this work is the realization of calculation of the main csientometrical indexes (quotation index and Hirsch index) on the base of data from RSCI and SCOPUS under the conditions of noncomplete determination.

To solve the presented problem there was developed the approach allowing make the analyses of the main scientometrical indexes such as quotation index, Hirsch index and others on the base of data from quotation systems under the conditions of noncomplete determination. The main components of developed approach and their relations are shown through IDEF0-diagram of the first level on picture 1.



Picture 1 – IDEF0-first level diagram.

Let's consider the analyses of authors' publication activity on the base of RSCI and SCOPUS quotation systems.

At the stage of data collection and preparing is performed the review of html-pages with the list of publications and quotations from RSCI and SCOPUS quotation systems.

By building of aggregative list of publications it's being formed the list of author's publications on the base of data from RSCI and SCOPUS quotation systems in which there are no duplicated publications. The base of building algorithm of aggregative list of publications makes shingles algorithm.

The stage of data collection and preparing and also the stage of aggregative list of publications building are presented more detailed in work [1].

To calculate the total number of quotations on each publication there are used data received at the stage of building of aggregative list of publications precisely the number of quotations in RSCI system, the number of quotations in SCOPUS system and also the number of found duplicated publications in SCOPUS system.

The base of calculation algorithm of total number of quotations under the conditions of noncomplete determination makes mathematical apparatus of fuzzy decision trees [6]. This mathematical apparatus combines the advantages of decision trees and fuzzy logic.

There were detached 2 target classes: «small proportion of intersectional quotations» (negative result), «big proportion of intersectional quotations» (positive result).

The belonging to the target class for new recording can be found in the following way:

$$\delta_j = \frac{\sum_l \sum_k P_k^l \cdot \mu_l(D_j) \cdot x_k}{\sum_l (\mu_l(D_j) \cdot \sum_k P_k^l}\,,$$

where P_k^l - coefficient reflecting the correlation of sheet examples *l* for the meaning of the whole class *k*, $\mu_l(D_j)$ - belonging level example *j* to the knot *l*, x_k - target class meaning belonging *k* to the positive result.

The calculation of total number of quotations is made in the following way:

$$S_i = K_i^{SCOPUS} + K_i^{RISC} - \min(K_i^{SCOPUS}; K_i^{RISC}) \cdot \delta_i$$

where K_i^{SCOPUS} - the number of quotations in SCOPUS system for the current publication, K_i^{RISC} - the number of quotations in RSCI system for the current publication.

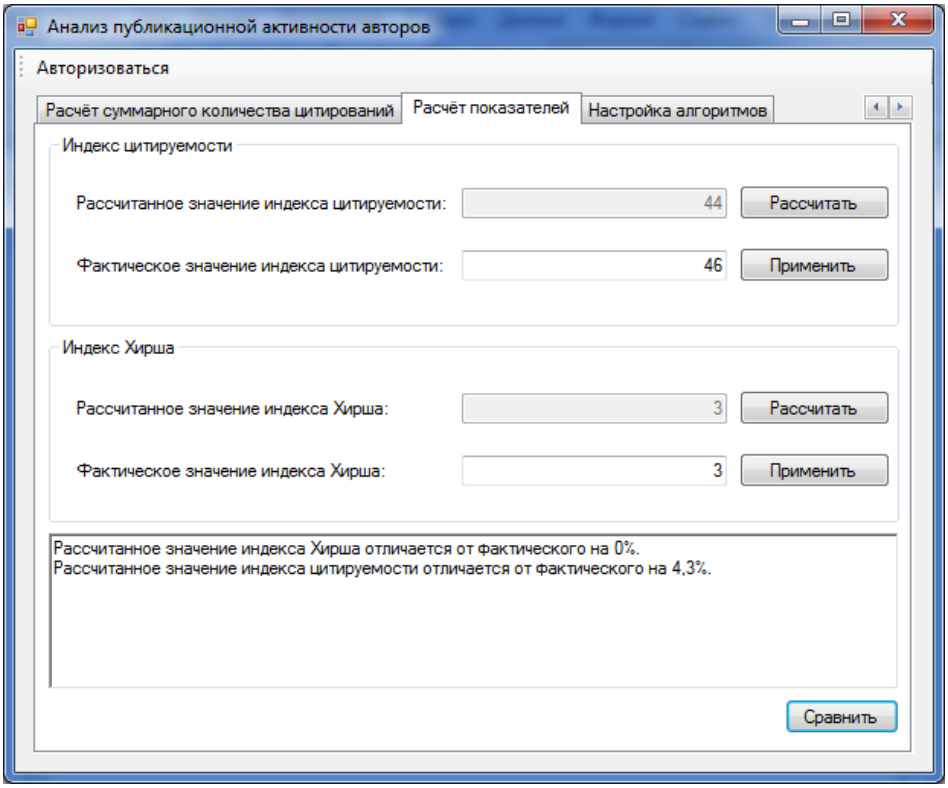
The citation index is defined as the sum of quotations calculated at previous stage in all publications. Hirsch index is calculated by formula presented in work [4] on the base of total number of quotations calculated at previous stage.

The results of citation index and Hirsch index calculation for one of the authors included in test set are presented in table 1 and on picture 2.

Table 1

The results of citation index and Hirsch index calculation

The name of index	Index value in RSCI	Index value in SCOPUS	Calculated index value	Index real value
Citation index	35	13	44	46
Hirsch index	2	2	3	3



Picture 2 – The screenshot of the program at the stage of the main scientometrical indexes calculation.

Conclusion

The results of citation index and Hirsch index on the base of developed approach allowed make following conclusions: Hirsch index calculated value matches with real value; calculated value of citation index differs slightly from the real one.

As it can be seen from the above it's possible to talk about acceptable results of developed approach and the possibility of its further usage for citation index and Hirsch index calculation.

References:

1. Web of Science: online subscription-based scientific citation indexing service. Available: http://isiknowledge.com (Accessed: 15.02.2017).
2. Scopus: citation database of peer-reviewed literature. Available: http://www.scopus.com/ (Accessed: 15.02.2017).
3. eLIBRARY.RU: nauchnaja jelektronnaja biblioteka. Available: http://elibrary.ru (Accessed: 15.02.2017).
4. Kotsemir MN (2012) Publication Activity of Russian Researches in Leading International Scientific Journals. Acta naturae. V. 4 N. 2 (13), p. 15-35.
5. Garfield E, Paris SW, Stock WG (2006) HistCite: a software tool for informetric analysis of citation linkage. Infometrics. N. 57, p. 391-400.
6. Baneyx A (2008) "Publish or Perish" as citation metrics used to analyze scientific output in the humanities: international case studies in economics, geography, social sciences, philosophy, and history. Archivum Immunologiae Et Therapiae Experimentalis. V. 56, N. 6, p. 363-371.
7. Kiduk Y, Lokman IM (2006). Citation Analysis: A Comparison of Google Scholar, Scopus, and Web of Science. Proceedings of the American Society for Information Science and Technology. V. 43, N. 1, p. 1-15.
8. Krylov IB, Boldyrev PA (2015) Several characteristics of existing automated systems according to survey of russian scientists publishing activity. Theoretical & Applied Science. N. 5 (25), p. 6–9.
9. Boldyrev PA, Krylov IB (2016) Razrabotka agregirujushhej sistemy analiza publikacionnoj aktivnosti uchjonyh na osnove mezhdunarodnyh i rossijskoj sistem citirovanija v uslovijah ogranichennogo dostupa. Vserossijskaja nauchno-metodicheskaja konferencija "Universitetskij kompleks kak regional'nyj centr obrazovanija, nauki i kul'tury". Orenburg, p. 2602-2608.
10. Hirsch JE (2005) «An index to quantify an individual's scientific research output». Proceedings of the National Academy of Sciences. N. 102, p. 16569-16572.