Scientometric Analysis of Research Papers from Serbia in the Field of Computer Science

Jelena Cvetinović*, Uroš Romić*, Miloš Pavković*, Jelica Protić*
* University of Belgrade, School of Electrical Engineering, Belgrade, Serbia
cj213265m@student.etf.bg.ac.rs, uros.romic@etf.bg.ac.rs, milos_pavkovic@yahoo.com, jelica.protic@etf.bg.ac.rs

***Abstract*—In this paper we examine the scientific production of Serbian scientists in the area of computing, published between year 2008 and year 2020, extracted from the Web of Science (WoS) database. Our dataset includes papers published in journals with impact factor, which are classified in at least one out of seven WoS categories with the prefix "Computer Science". The work includes identification of unique authors' and institutions' identity from the dataset, as well as creation of a database used for bibliometric and coauthorship analysis. As a result, we have created coauthorship network using Gephi, identified the leading authors and institutions, as well as the most popular journals, WoS categories and collaborators of Serbian authors in the field of computer science for the analyzed period.**

I. INTRODUCTION

During the last decade, a considerable progress emerged in the information technology sector in Serbia. As a result of this trend, the growing interest of high school graduates in enrolling to the study programs in the field of computing [1] resulted in insufficient number of teaching staff, especially since the positions in industry became more attractive for young experts. The future of this sector depends on the quality of education and sufficient number of teaching staff with expert knowledge appropriately distributed over sub-disciplines of computer science. However, the ultimate condition for their promotion at universities are scientific papers with impact factor. For this reason, analysis of scientific publications in computer science would bring additional insight into the potentials of education and research in this field. The scientific production of individuals and research groups is the crucial factor in accreditation of higher education institutions, especially doctoral studies, as well as projects approval and evaluation.

The main subject of this paper is an analysis of scientific production of authors from Serbia in seven categories with the prefix "Computer Science", based on their journal papers indexed in the Web of Science (WoS). The aim of this analysis is to determine and calculate certain bibliometric indicators and construct the social network of coauthorship in order to reveal collaboration patterns between Serbian computer scientists, as well as their collaborations with partners from abroad. Another goal is to identify the strength in certain subfields that can encourage young researchers to choose their fields of interest and improve cooperation between research groups and individuals.

.

II. RELATED WORK

Multiple bibliometric and scientometric research results that focus on the broad area of computer science were published in the open literature. The source of data about published papers has strong influence on the results. Computer scientists sometimes intuitively feel the assessment criteria, usually grouped together with mathematics and/or electrical engineering, to be unfair to computer scientists [2]. The authors of the study [2] state that on average 66% of a computer scientist's published work is not accounted for in Web of Science (WoS). In the paper [3] authors examined citations to over 195,000 conference papers and 108,000 journal papers in CS, concluding that overall citation rates in conferences are not higher than in journals, although there are some top CS conferences with the highest average paper citation rate of any publication type. However, in [2] authors consider WoS to be the most traditional citation-indexing service and they confirm that it became standard in most bibliometric research. Therefore, WoS was primary data source for our analysis, and papers published in scientific journals were considered.

On the global level, a recent study [4] examined the research output of countries in CS worldwide, over the past 25 years (1995–2019). This paper shows an analysis for all journals indexed in the WoS in the seven categories of computer science research, which is the same approach used in our study. The paper [4] shows that United States is the leading country in the computer science field, and that some other countries, such as the United Kingdom, China, Canada and Germany, obtain very high positions in the ranking.

In Serbia, on the national level, the first wider bibliometric analysis for all scientific fields was presented in the paper published in 2014 [5]. It included 14,293 articles having all authors from Serbia, published in the period 2006–2012, indexed in the SCI-EXPANDED database. Distribution of articles in the Web of Science categories, journals, scientific research institutions and researchers were analyzed, but international collaboration was not considered. This analysis identified Stevic, S. followed by Gutman I. as the most productive authors, while the WoS category with the highest number of papers was "General and internal medicine", partially due to 705 papers published in *Vojnosanitetski pregled*, the journal located in Serbia which entered SCI-EXPANDED in 2008.

In order to compare the results from [5] with the results for the specific area of computer science, we have performed a preliminary analysis for the period 2006–

2013 presented in [6]. It included all papers with at least one author from Serbia, and all WoS categories with the prefix "Computer Science". This study identified Gutman I. and Ivanovic M. as the most productive authors, and the highest numbers of papers were published in journals MATCH and ComSIS, both located in Serbia. Data extraction proved to be prone to errors, due to various problems with affiliations, names and personal identity [7] so it required preprocessing and data cleansing.

Another paper which concentrates on Serbian authors from our institution, University of Belgrade - School of Electrical Engineering (UB-SEE), used data about journal papers from the institutional database over the period from 2000 to 2017, in order to create, examine and visualize the coauthorship network [8].

As computer science became more popular, several studies from different countries were published that analyzed scientific output in the CS field on the national level. A study that focuses on Mexican scientific production in CS [9] identifies the most productive institutions, shows country-level and institution-level collaboration networks, as well as keyword density diagrams. An analysis of Brazilian CS research [10] starts from the list of 406 scientists who participate in the Researcher Fellowship Program. Besides their publication volume and citation count, distributed over 23 CS areas in the period from 2001 to 2011, their career length and mentorship are also analyzed. As an example from Europe, the cluster analysis methodology has been applied to the Spanish public universities and their academic staff in the computer science area [11]. Four different clusters of universities and six different clusters of academic staff were identified, based on characterization of their research activities. All these papers served as a motivation and starting point for our analysis presented in this paper.

III. RESEARCH QUESTIONS

In this research, our aim was not only to analyze the data about scientific production that we have retrieved from the WoS database, but also to identify and describe the problems of data retrieval process itself.

Accuracy of the data retrieval that we have performed using Web of Science queries in order to provide data set for our analysis is particularly important and the results can be compared to those of alternative institutional system which provides data about published papers input by the authors. Results from these two data sources were compared in order to draw conclusions about the differences and to get more precise results.

The quality of data highly influences the results, so it has to be sufficient for further analysis. Therefore, we were interested to find out what can be done in the process of data cleansing in order to improve the data quality and consistency.

Regarding the retrieved data and data processing results, our goal was to answer who are the most productive scientists from Serbia who have published papers in journals with impact factor, indexed in Web of Science categories with the prefix Computer Science. We analyzed if there are significant differences in ranking if we use the total number of published papers compared to the fractional counting, where the number of coauthors is taken into account.

We analyzed the productivity of institutions of higher education and research in Serbia in the field of Computer Science, trying to answer which are the most productive institutions based on publications indexed in Web of Science. Besides productivity, we also analyzed the cooperation of Serbian computer scientists with the institutions and individuals from abroad that have resulted in coauthorship on published papers.

Impact of journals is of particular importance to authors, so we analyzed which are the journals indexed in Web of Science categories with the highest numbers of papers published by Serbian authors, and how many of these journals are published in Serbia.

The sub-areas of computer science were also part of our analysis. Identifying what are the sub-areas with the highest numbers of published papers written by Serbian authors is very important for policy makers and young researchers deciding on their primary field of interest.

Finally, we wanted to compare how the results obtained in this analysis correspond with the previous bibliometric analysis for the period 2006-2013.

IV. METHODOLOGY

We have extracted data from the Web of Science in csv form, using queries to select records for papers that have at least one author with the affiliation of an institution in Serbia, in the period 2008-2020 for journal papers indexed in at least one Web of Science category beginning with "Computer Science" (e.g. Computer Science, Artificial Intelligence; Computer Science, Cybernetics; Computer Science, Hardware & Architecture; Computer Science, Information Systems; Computer Science, Interdisciplinary Applications; Computer Science, Software Engineering; Computer Science, Theory & Methods). The csv file was imported to Microsoft Excel for visual inspection.

The following WoS fields [12] were identified as the most important for the purpose of this work:

- AF - Author Full Name
- TI - Document Title
- PY - Year Published
- C1 - Author Address
- WC - Web of Science Categories
- SN - International Standard Serial Number (ISSN)
- EM - E-mail Address
- SC - Research Areas

The most important data field for connecting authors with institutions and countries is the field C1, which contains the data about addresses of authors and their institutions, for example:

„[Geler, Zoltan] Univ Novi Sad, Fac Philosophy, Dept Media Studies, Dr ZoranaDindica 2, Novi Sad 21000, Serbia; [Kurbalija, Vladimir; Ivanovic, Mirjana; Radovanovic, Milos] Univ Novi Sad, Fac Sci, Dept Math & Informat, Trg D Obradovica 4, Novi Sad 21000, Serbia“

Visual inspection has shown that the order of data in this field is not unified, and that some records are incomplete.

For example, there are records which contain more than one institution per author (*„[Dimitrov, Darko] Hsch Tech & Wirtschaft Berlin, Berlin, Germany; Fac Informat*

Studies, Novo Mesto, Slovenia; [Milosavljevic, Nikola] Univ Nis, Fac Sci & Math, Nish, Serbia"), records which contain information about institution but without author (*„Univ Novi Sad, Fac Sci, Novi Sad, Serbia; [Nanopoulos, Alexandros] Katholische Univ Eichstatt Ingolstadt, D-85072 Eichstatt, Germany"*), and also records with information about multiple institutions but with no authors at all (*„Univ Nis, Fac Sci & Math, Nish 18000, Serbia; Univ Nis, Fac Econ, Nish 18000, Serbia"*).

Most of the records in C1 field conform to the following pattern: *[Author's Name] Name of the institution {, additional data}, city, county*.

However, nine different cases of given data and ordering in this field were identified, for example: there is additional data about the city; name of the city and name of the country were swapped, different names were used for the same city (*Belgrade, Zemun, Novi Beograd, N Belgrade, New Belgrade, Beograd*), differences in additional data for the same institution (address), errors in the names of the institutions etc.

Another important field for data processing is AF, which contains full names of authors separated by semi colon. Several inconsistencies were identified in this field, such as wrong spelling of names and different usage of initials. In some cases, author's email address (from EM field) was used to correctly identify authors.

Some inconsistencies required manual corrections of imported data, and after that the data was ready for processing by the parser, written in Java.

The data from Excel file was imported into parser using Apache POI 3.17 library. After iterating through Excel sheet, the data was processed and imported into *MySql* database using several Java methods, which performed extraction of WoS categories and research areas, extraction of institutions based on several different patterns for recognizing institution data, correction of author names, email addresses, and institution names, and, finally, connection of authors, papers and institutions using database relations. To connect Java analytical program with the database *mysql-connector-java-5.1.49* driver was used.

After creating independent tables *papers*, *institutions*, *authors*, *wos_categories* and *research_areas*, we have created *Many-to-many* tables *authors_institutions*, *papers_authors_institutions*, *papers_woscategories*, and *papers_research_areas*. To inspect the data obtained after data cleansing, *phpMyAdmin* was used.

V. Solution/Discussion

After data processing and creation of the database, we have identified 2289 records about papers, with the total of 3920 authors coming from 2113 different institutions.

Considering productivity of individual scientists, we have performed total and fractional counting, where we assigned 1/n fraction to each author for the paper with n coauthors.

From this data set, we identified the most productive institutions, both for fractional counting and total number of papers in the dataset (Table I). For fractional counting, we valued each coauthor from the same institution as 1/n, with n being the total number of coauthors. For example, if one paper had 2 coauthors from the same institution and total number of coauthors is 5, that paper was valued as 2/5 for the given institution.

TABLE I.
THE MOST PRODUCTIVE INSTITUTIONS BASED ON TOTAL AND FRACTIONAL COUNTING OF PAPERS IN THE DATASET

Institution	Fractional counting	Number of papers
Univ. Belgrade	562.14	822
Univ. Novi Sad	405.88	603
Univ. Nis	279.27	407
Univ. Kragujevac	149.59	279
Serbian Acad Arts & Sci	78.92	192
Univ. Pristina	23.63	45
Univ. Singidunum	22.25	64
Univ. State Novi Pazar	16.10	50

Some differences can be noticed in ordering of institutions by fractional counting, compared to the ordering by total number of papers, because some institutions have fewer authors per paper on average. For example, University of Pristina has better rank in fractional counting than in total number of papers. However, for the leading three universities, the relative difference in productivity is similar in total number of papers, compared to the fractional counting values.

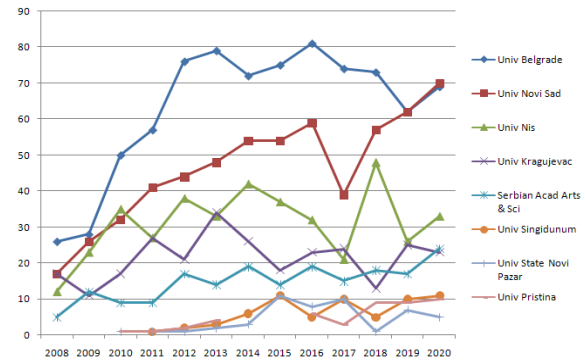

Figure 1. Total number of papers per institution per year

Total number of papers per institution per year is shown in Figure 1. The increase in numbers of papers over this period is most visible for the University of Belgrade and University of Novi Sad, with latter taking the leading position recently.

We have also identified leading authors, with the highest numbers of published papers based on fractional counting and total number of papers (Table II), as well as leading WoS categories (Table III). It can be noticed that there are significant differences in the case of ranking the leading authors by total number of papers, compared to ranking by the values resulting from fractional counting.

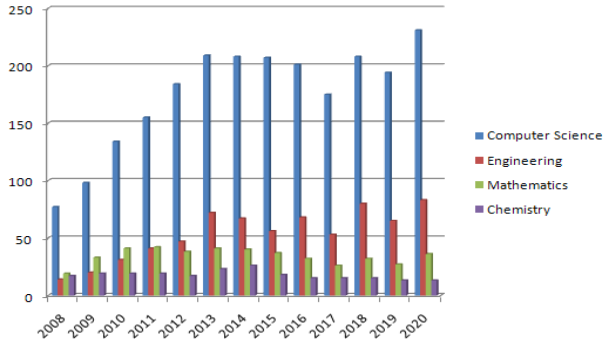

Figure 2. The most frequent research areas

The most frequent research areas in which the journals from our data set were classified over the years by the number of papers classified in each area are shown in Figure 2. It shows that overlapping of CS with engineering, mathematics and chemistry is strong and proves its interdisciplinary nature.

Leading journals by the total number of papers are presented in Table IV. Impact Factor (IF) published for the year 2020 for each journal is also shown. Journals MATCH and COMSIS, which have significantly larger number of papers compared to other leading journals, are also the only journals published in Serbia which are classified in Computer Science and have Impact Factor.

TABLE II.
THE MOST PRODUCTIVE AUTHORS BASED ON FRACTIONAL COUNTING OF PAPERS

Author	Fractional counting	Number of papers
Gutman Ivan	35.03	104
Ivanovic Mirjana	16.38	61
Stevanovic Dragan	15.10	23
Peric Zoran H.	12.68	40
Zunic Jovisa	12.40	25
Pap Endre	10.96	27
Tepavcevic Andreja	9.72	26
Brodic Darko	9.08	22
Filipovic Nenad D.	8.09	38
Ciric Miroslav	8.00	25

It should be noted that, during the initial data analysis, we have received additional information about some retracted papers and papers in the procedure of retraction. Total of 8 papers have been identified and removed from the data set for final analysis and results due to retraction.

All of the journals included in the analysis have been checked online and none of them has been recognized as a potentially predatory journal, according to Beall's list. One of the popular journals, *Mathematical and computer modeling*, was discontinued in 2014, so the last Impact Factor published for this journal is for year 2015.

TABLE III.
LEADING WOS CATEGORIES BASED ON NUMBER OF PAPERS

WOS category	Number of papers
Computer Science, Interdisciplinary Applications	814
Computer Science, Artificial Intelligence	651
Computer Science, Information Systems	545
Computer Science, Software Engineering	391
Computer Science, Theory & Methods	368
Computer Science, Hardware & Architecture	141
Computer Science, Cybernetics	41

TABLE IV.
LEADING JOURNALS BASED ON NUMBER OF PAPERS (WITH THEIR LATEST WOS IMPACT FACTOR)

Journal Title	Number of papers
MATCH: communications in mathematical and in computer chemistry (IF 2020 = 2.497)	158
Computer science and information systems / COMSIS (IF 2020 = 1.167)	112
Expert systems with applications (IF 2020 = 6.954)	73
Computer applications in engineering education (IF 2020 = 1.532)	60
Advances in electrical and computer engineering (IF 2020 = 1.221)	58
Fuzzy sets and systems (IF 2020 = 3.343)	57
Mathematical and computer modelling (discontinued in 2014) (IF 2015 = 1.366)	44
Information sciences (IF 2020 = 6.795)	43
Journal of multiple-valued logic and soft computing (IF 2020 = 0.861)	42
IEEE Access (IF 2020 = 3.367)	37

If we compare the list of institutions (Table I) with the list of scientists (Table II), we can notice that there is one scientist form this table who is affiliated with the University of Belgrade - D. Brodić from the Technical Faculty in Bor. There are two scientists from the University of Kragujevac (I. Gutman and N. Filipović), two scientists from the University of Novi Sad (M. Ivanović and A.Tepavčević), two scientists from the University of Niš (Z. Perić and M. Ćirić), two scientists from the institute MISANU (D. Stevanović and J. Žunić) and one scientist from Singidunum University (E. Pap).

Collaborations of four leading institutions from Serbia is shown in Table V. Besides collaborations between institutions and authors from Serbia, which are understandably the most numerous, it can be noticed that the most collaborations are made with scientists from institutions based in the USA and Germany. However, University of Kragujevac and University of Niš also have relatively high number of collaborations with China and Iran, compared to the University of Belgrade and University of Novi Sad.

TABLE V.
COLLABORATIONS OF FOUR LEADING INSTITUTIONS FROM SERBIA

University of Belgrade		University of Novi Sad	
country	No. of collab.	country	No. of collab.
Serbia	94	Serbia	61
USA	42	USA	25
England	28	Germany	22
Spain	27	England	20
Germany	21	Italy	14
France	19	Spain	12
Italy	14	France	11
Greece	10	China	10
University of Kragujevac		University of Niš	
country	No. of collab	country	No. of collab
Serbia	34	Serbia	43
China	27	Germany	14
USA	23	China	11
Iran	14	Iran	7
Italy	11	France	6
India	9	South Korea	6
Germany	8	Greece	5
Greece	7	USA	5

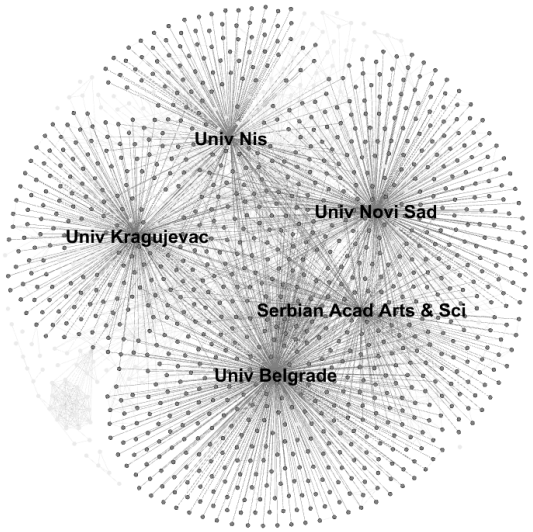

Figure 3. Collaboration network for leading institutions in Serbia

Regarding collaborations between individual authors, leading individual collaborators are Filipović Nenad with 411 collaborations and Gutman Ivan with 338 collaborations.

Collaboration data was processed using Gephi, an open-source software for network analysis and visualization, written in Java. Data was imported with institutions or authors as nodes of the graph, and papers they collaborated on as graph edges. Undirected and unweighted graph was used i.e. multiple collaborations between authors were considered as one, and graphical representation was formed using Fruchterman Reingold layout. Resulting diagram for collaborations between institutions, with leading institutions highlighted, is shown in Figure 3. Collaboration network for authors with most collaborations is shown in Figure 4.

Besides just detecting that collaboration between institutions existed, the scale of collaboration was also measured, by the number of common papers published. For example, if there was one or more authors of the same paper from two or more institutions, that was counted as one collaboration between each of authors' institutions i.e. each pair of those institutions have one paper as a result of collaboration. Table VI. shows how many papers were published as a result of collaboration between institutions.

Table VII shows the leading tree institutions with the highest numbers of collaborations (shown in Figure 3.) and, for each of them, three institutions that had the highest number of collaborations with them.

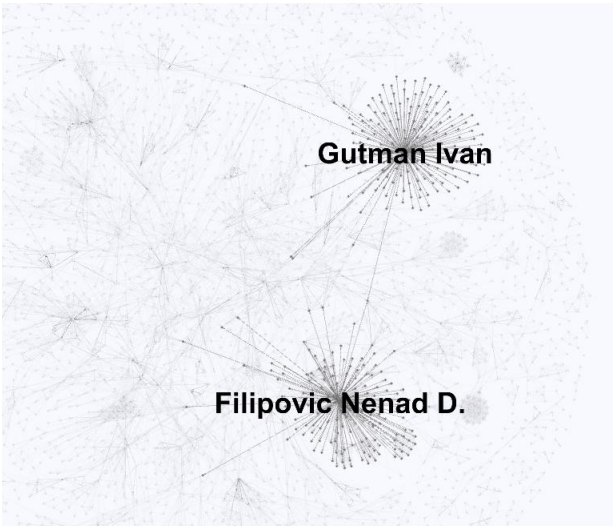

Figure 4. Collaboration network for authors with most collaborations

TABLE VI.
SCALE OF COLLABORATION BETWEEN INSTITUTIONS BY THE NUMBER OF PAPERS PUBLISHED

		No. of papers
Univ. Belgrade	Serbian Acad Arts & Sci	47
Univ. Belgrade	Univ. Novi Sad	37
Univ Kragujevac	BioIRC	29
Univ Belgrade	Univ Kragujevac	29
Univ Belgrade	Univ Nis	21
Univ Kragujevac	Univ State Novi Pazar	21
Univ Nis	Univ Pristina	20
Univ Novi Sad	Serbian Acad Arts & Sci	19
Univ Nis	Univ Kragujevac	17
Serbian Acad Arts & Sci	Univ Nis	16

TABLE VII.
SCALE OF COLLABORATION BETWEEN INSTITUTIONS BY THE NUMBER OF PAPERS PUBLISHED

		No. of papers
Univ. Belgrade (113)	Serbian Acad Arts & Sci	47
	Univ Novi Sad	37
	Univ Kragujevac	29
Serbian Acad Arts & Sci (82)	Univ. Belgrade	47
	Univ Novi Sad	19
	Univ Nis	16
Univ Kragujevac (79)	Univ. Belgrade	29
	BioIRC	29
	Univ State Novi Pazar	21
Univ Novi Sad (68)	Univ. Belgrade	37
	Serbian Acad Arts & Sci	19
	Univ Obuda	12
Univ Nis (58)	Univ. Belgrade	21
	Univ Pristina	20
	Univ Kragujevac	17

VI. CONCLUSIONS AND FUTURE RESEARCH

In this paper, the scientific production of Serbian scientists in the field of Computer Science was analyzed by collecting WoS data and processing it using Java parser. Most productive authors and institutions were identified and their collaborations with other institutions was analyzed and presented.

The data set covers the period from 2008 till 2020 and one of its most noticeable characteristics is that the total number of published papers per year increased almost three times during this period. This is certainly related to the global increase in scientific production in Serbia, as well as rising popularity of Computer Science.

During this research, the attention was paid to exclude the papers that are subject of retraction, and it was checked that none of the papers is published in predatory journals. By the number of papers published, the leading journals are the only two Serbian journals classified in the field of Computer Science which have IF and their editors-in-chief I. Gutman and M. Ivanovic remained the leading authors, as it was stated in the previous analysis [6]. The most popular WoS categories in Computer Science for Serbian authors are Interdisciplinary Applications and Artificial Intelligence, just like in the period 2006–2013 analyzed in [6], and the most popular journals for out CS authors remained the same, with addition of *Computer applications in engineering education* and *IEEE Access*.

Results of this analysis should be taken with caution because of numerous cases of inconsistent WoS data related to authors' affiliations. Advanced techniques of connecting authors and affiliations will be implemented in the future work, as well as alternative ways of checking connections between authors and institutions with more precision. Classifying as computer scientists all researchers who publish papers in journals that are listed in one of Computer Science WoS categories is not always correct, having in mind interdisciplinary nature of this field. In addition, it should be noted that our analysis is limited to productivity, so it does not take into account widely accepted research indicators such as h-index or g-index, identification of highly cited authors and papers etc.

The future work will also include additional citation analysis, analysis of keywords of papers and analysis based on alternative classification of scientific fields. Also, the future research will consider alternative data sources, as well as the correlation between the increase in scientific production and the increase in number of teaching staff and students in the field of computer science, which are of crucial interest for the growth of working force in this field..

REFERENCES

[1] M. Savić, M. Ivanović, I. Luković, B. Delibašić, J. Protić,. and D. Janković, "Students' preferences in selection of computer science and informatics studies: A comprehensive empirical case study," *Computer Science and Information Systems*, vol. 18, iss. 1, pp.251-283, 2021.

[2] J. Wainer, S. Goldenstein, and C. Billa, C., "Invisible work in standard bibliometric evaluation of computer science," *Communications of the ACM*, vol. 54, no. 5, pp.141-146, May 2011.

[3] G. Vrettas and M. Sanderson, "Conferences versus journals in computer science," *Journal of the Association for Information Science and Technology*, vol. 66, no. 12, pp.2674-2684, Dec 2015.

[4] G. Zurita, J.M. Merigó, V. Lobos-Ossandón, and C. Mulet-Forteza, "Leading countries in computer science: A bibliometric overview, " Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp.1957-1970, Jan 2021.

[5] D. Ivanović and J. S. Ho, "Independent publications from Serbia in the Science Citation Index Expanded: a bibliometric analysis," *Scientometrics*, vol. 101, iss. 1, pp. 603-622, 2014.

[6] M. Pavković and J. Protić, "An Analysis of Scientific Publications from Serbia: The Case of Computer Science," *Proceedings of the15th International Society of Scientometrics and Informetrics Conference*, pp. 473 - 478, ISSI, Istanbul, Turkey, June 2015.

[7] I. Mitrovic, J. Protić, "Problems with affiliations, names and personal identity in the process of evaluating higher education institutions," *Proc. of EDULEARN14, Barcelona*, pp. 2524-2533, 2014.

[8] J. Stojković, M. Mišić and J. Protić, "Collaboration network analysis of scientific production at UB-SEE," XXVII Telfor 2019, Telecommunication Society, Belgrade, Nov, 2019.

[9] A. Uddin, V.K. Singh, D. Pinto and I. Olmos, 2015. "Scientometric mapping of computer science research in Mexico," *Scientometrics*, vol. 105, no. 1, pp.97-114, Oct 2015.

[10] H. Lima, T.H. Silva, M.M. Moro, R.L. Santos, W. Meira, and A.H. Laender, "Assessing the profile of top Brazilian computer science researchers," *Scientometrics*, vol. 103, no. 3, pp.879-896, Jun 2015.

[11] A. Ibáñez, P. Larrañaga and C. Bielza, 2013. "Cluster methods for assessing research performance: exploring Spanish computer science," *Scientometrics*, vol. 97, no. 3, pp.571-600, Dec 2013.

[12] Web of Science Core Collection Field Tags, https://images.webofknowledge.com/images/help/WOS/hs_wos_fieldtags.html