

Open Source Software for Arabic Citation Engine: Issues and Challenges

Saleh Alzeheimi, Akram M. Zeki, Adamu I Abubakar

Abstract— Recently, there are various software for citation index such as Scopus, Google Scholar, Journal of Citation Report (JCR) and others. These software have some disadvantages such as not providing support to the literature which is written in Arabic and built as commercial and closed software. This study aims to explore Arabic initiatives in the development of OSS. An Arabic citation engine is an alternative to the existing software. It also aims to investigate the ability of Systematic Analysis for Arabic Citation Index (SAACI)

Index Terms— Bibliometrics management software, Open source software, Citation Index, Systematic Analysis for Arabic Citation reference (SAACI).

XXXII. INTRODUCTION

BIBLIOMETRICS is one of important scientific research tools for analyzing the characteristics of literature in specific subjects. This tool helps researchers and institutions to know relatives or separated subjects, knowledge sharing, the core of resources, impact factor of journals, high contribution among authors and other systematic analysis. According to Qasim [1], There has been increasing in number of journal requests to find the appropriate scientific tools to choose and evaluate the printed journals by librarians. In view of this importance, it has developed many systems for foreign mechanism which analyze intellectual foreign heritage, statistical indications and measurements of bibliometric. The most popular systems include, Scopus and Journal of Citation Report (JCR).

With the availability of many studies the mechanism of these systems is discussed, whether through the use of analysis of a particular intellectual heritage or through the evaluation of these systems, and the extent of the commitment for measurements and bibliometric, laws. But Arabic Studies (which means an analysis of Arab intellectual heritage) have not benefitted much from this system, as it does not support sources and citations written in Arabic language. This forces researchers and those who focus on bibliometric studies and analysis of characteristics of a particular intellectual heritage, to use the manual methods of analysis and application of bibliometric laws. With the existence of this gap, this study will attempt to explore Arabic initiatives in the development of OSS Arabic citation engine as an alternative to the existing software. The study will focus on answering two research questions:

- What are the current Arab initiatives in the development of open source software as the Arabic citation engine?
- What are the technical aspects and analysis for Systematic Analysis for Arabic Citation Index (SAACI)?

Hence, the answer to these questions will allow for the investigation of the ability and readiness of utilizing (SAACI).

XXXIII. RELATED WORKS

Among the studies, which deal with the characteristics of the Arab intellectual heritage, is the study by Timraz [2] who identified the characteristics of intellectual heritage that can be used by the Arabic researchers in the area of engineering in the Kingdom of Saudi Arabia. In the same year, Al Dausari [3] used in his study the bibliometric curriculum, entitled “An Information Communication among Arabia Researchers in Pure Sciences”. In 1993, the study of Turkustani [4] use biometric law to explore the characteristic of intellectual heritage in the field of libraries and information In same year, Al-haddad [5] used a bibliometric analysis to study the citations in the Alam Alkotob magazine. In year 1994 Qandil [6] focused on analysing citations in the *Resalat Almaktabah* magazine. In year 2002 the study of “Electronic Publishing in Ten Years 1990-1999 Bibliometric study” by both Hamdi and Ghneim [7] aimed to identify the main features of an intellectual heritage. In 2003, Abdul Sattar [8] addressed in his study the electronic source for information in the area of library and information by analyzing citation references with Internet sources in Arabic periodicals. In 2003 Zayed [9] prepared a study on electronic source for distance citation references. She analysed theses, documentation and information in Cairo during the period of 1998-2003. In 2004, the study by Alumer [10] described characteristics of an intellectual heritage in the field of information technology through analysis of citations in Arab periodicals.

By tracing the tools used by the researchers in the above-mentioned mentioned studies, it is clear to the researcher, that these studies and others used manual methods to analyse Arabic intellectual heritage, where some of these studies used tables only to easily calculate averages and other simple small processes.

However, the study by Al-zeheimi [11] is the first Arabic study that developed the automated systems for open source, and was built on the web based for being most bibliometric measurements.

The second part of the literature is on Arabic studies that address applications for automated systems in the area of bibliometric studies. However, most of these studies are foreign studies except for the study by Kelow [12] and by Al-Najjar [13] who addressed the automated foreign

programs such as Scopus and JCR, but do not highlight the Arabic system, which was built and developed by Al-zeheimi [11,14], through his study of Omani intellectual heritage in the medical field. Al-zeheimi [14] has indicated this system and its characteristics in his paper presented at the conference of specialized libraries, in Arab Gulf in Kuwait in 2008.

To sum it up, it is observed that most of the previous studies used manual methods in Arabic intellectual heritage study, due to the weakness of potential automated systems that has been developed, such as Scopus and JCR in supporting Arabic publications and analysing their characteristics, as well as lack of Arabic intellectual heritage in electronic databases, and the rules of world information, compared to foreign intellectual heritage. Furthermore, there is no Arabic open source software built in web based, except the system for Systematic Analysis for Arabic Citation Index (SAACI) which was developed by Al-zeheimi [11,14], and used in his study on the Omani intellectual medical heritage.

Thus, the current study will address in detail the Systematic Analysis for Arabic Citation Index (SAACI) and its appropriateness in massive usage by specialists in this area, and the extent of its compatibility with the requirements of the Arab intellectual heritage.

XXXIV. METHODOLOGY

The study adopts an analytical method, to study the characteristics of the system for Arabs in the area of bibliometric studies, as well as addresses foreign systems, such as Scopus and JCR, in order to take advantage of the measurements, reports and statistics that are presented. Finally, the study highlights an Arabic open source system (systematic Analysis for Arabic citation index - SAAIC).

A. Systematic analytic tools

The diffusion of systematic analysis tools or applications is increasing rapidly nowadays. The most popular of these tools are Scopus, Journal of Citation Report (JCR) and Google scholar.

1) Journal Citation Report (JCR)

Journal Citation Reports offers a systematic, objective means to critically evaluate the world's leading journals, with quantifiable, statistical information based on citation data. By compiling these articles' cited references, JCR helps to measure research influence and impact at the journal and category levels, and show the relationship between citing and cited journals. This is available in Science and Social Sciences editions (<http://thomsonreuters.com/journal-citation-reports>). JCR provides some bibliometric features such as Impact Factor, Immediacy Index, Total Cites, Total Articles, Cited Half-Life or Journal Title.

2) Scopus

As quoted from Wikipedia, cited from Scopus info; "*Scopus is a bibliographic database containing abstracts and citations for academic journal articles. It covers nearly 21,000 titles from over 5,000 publishers, of which 20,000 are peer-reviewed journals in the scientific, technical, and medical and social sciences (including arts and humanities)*". The sources informed that "*It is owned*

by Elsevier and is available online by subscription. Searches in Scopus incorporate searches of scientific web pages through Scirus, another Elsevier product as well as patent databases".

B. Arabic studies about systematic analysis tools:

There are many Arabic studies that use systematic analysis tools (Scopus, JCR and Google Scholar) to analyze their literature. Researchers use these tools for English publication, but when they want to analyze any Arabic publications, they usually use manual tools and a few of them use excel or access software. The main reason for that is because of the unavailability of bibliometrics analysis tools that support Arabic publications. This causes some weaknesses in Arabic literature in different phases as follows:

- Weaknesses in quantitative studies that address the characteristics of Arabic publications compared to studies that deal with foreign once.
- Citations from books, dissertations and theses, patents and technical reports are poorly covered.
- Science subject is poorly covered.
- Weaknesses in quantitative studies that address the characteristics of the Arab intellectual output compared to studies that deal with foreign publications.
- Limitation in analysis. Most Arabic studies analyzed bibliographic data, but they ignore the reference citation.
- Weaknesses of comprehensive studies which implement the main bibliometric elements because researchers find it difficult to apply some bibliometric laws without software.
- Thus, this study aims to explore the existing tools which may provide a solution to overcome the weaknesses.

XXXV. ARABIC INITIATIVES IN DEVELOPMENT OF CITATION ENGINE/TOOLS FOR ARABIC PUBLICATIONS

Bibliometric / systematic / webmetrics tools like Scopus or Google Scholar aim to provide reports based on the three most commonly used laws in bibliometrics: Lotka's law of scientific productivity, Bradford's law of scatter and Zipf's law of word occurrence. Such reports like number of papers, citations, average number of citations per paper and per author and per year as well as h-index, g-index, and other metrics are provided by these tools. Most researchers who study the literature which has been written in English use a software to get bibliometric reports. This paper explores studies that analyze the literature which has been written in Arabic. It found that most studies analyzed the data manually to get bibliometric reports. For example, Temars [2] studied the citations in engineering in Saudi Arabia. He used the traditional tools to calculate the citations. In the library field, most researchers use excel sheet to get reports and graphs [1-10].

Based on these findings, the current study will analyse only an Arabic system analysis, the systematic analysis for Arabic citation index (SAACI), and after that it will be evaluating in terms of possibility of further development, with what the researchers required, and the characteristics of Arabic intellectual heritage, also with the latest laws on bibliometrics.

A. *The idea of the software*

The idea of designing a open source software based on MySQL and PHP came during the study of Al-zeheimi [11, 14] for Oman intellectual heritage, in the field of Medical Sciences, where the researcher found most Arabic studies using traditional methods for analysis.

The great potential of PHP language as its open source in report design is that it is possible to be flexible and have the potential for data and retrieval articles or citations, through the design of searching, or have the potential to search by more than one field at the same time. In view of the requirement of bibliometrics from automated expert systems, the focus on extract statistical indicators which has scientific indicators, and based on the bibliography requirement and citation references, the idea of designing a Systematic Analysis for Arabic Citation Index has resulted in SAACI [11, 14]

B. *Advantages of the Software*

The current study examines SAACI to identify the features and technical advantages. The study also analyzes the available elements in the system and come up with the following:

- 1- The open source software (OSS), allows source files modification and development, and allow developers to participate in the solution of the problems of the development program.
- 2- The ability to analyse the various types of materials such as periodicals, articles, books, University theses and manuscripts.
- 3- Ease of use; one of the advantages of the program is easy to move between user interface and screen, through the presence of hyperlinking elements such as main screen or articles, research or statistics and others.
- 4- Working in a network environment; the program works through the Internet, or through a network in the local area.
- 5- Multiple Entry where the design has several levels for accessing the program and to deal with it. There is the power for the director, which enables him to deal with all functions of the program. Also the Supervisor is entitled to know some of the statistics access reports and so on.
- 6- Linking the tables; the program has automated features that link tables, making it easy for the researcher to process data entry, for example after data entry for a particular article, the program provides hyperlink to enter citation reference contained in the article on the same screen.
- 7- Easy to search for articles: it is possible to use the program to search for articles in all fields, or to search for citations. The program provides method of linkage between fields, which can possibly specify a research process for accurate results.
- 8- Automatic construction of the lists of authors or publishers, through appropriate fields. This advantage saves effort, time and accuracy for the researcher in the process of entering data related to authors or publishers or topics. It is also easy, by clicking on the name or topic that appears automatically in the lists. Therefore, the program will show the users the process of entering author's citations and references of the article if the author had been named for the articles, in other words, the article had been entered earlier in the program, in the sense that the program is working to provide lists of the authors, topics and publishers.
- 9- Auto Update statistics; when entering data for a new article or an author or publisher or topic, the system will automatically update all the statistics that have been identified in the program. Upon completion of the process of entering the new article, for example; when entering a particular article under specific periodical year, the program automatically adds an additional number to the total articles entered, and adds an additional number to the total articles published based on the years, and adds an additional number to the total periodical articles and so on, which means that only the researcher can enter data or articles or citations at once, and get comprehensive statistical data.
- 10- Prepare the reports automatically. Reports that the program makes are divided automatically into two sections; the first section focuses on analysis of articles while the second section focuses on analysis of citations

1) *C. Outputs of the Software*

The program consists of five main elements, namely Articles, Citations, Reports, Management and Lists. Figure 1 shows the five main elements. Each main element contains sub-fields. For example, for Articles you can add full texts or just bibliographic data. Also, you can search in different fields. Reports are the other main element in this software which contains statistics about authors in articles, authors in citations, average of citations in each article, average number of authors in each article and citation. Reports also contain statistics about the characters of authors either in articles or citations. From this element, we can know the authors who have in their cited articles, the types of resources cited and other important reports which are authors who have been most cited in their articles.

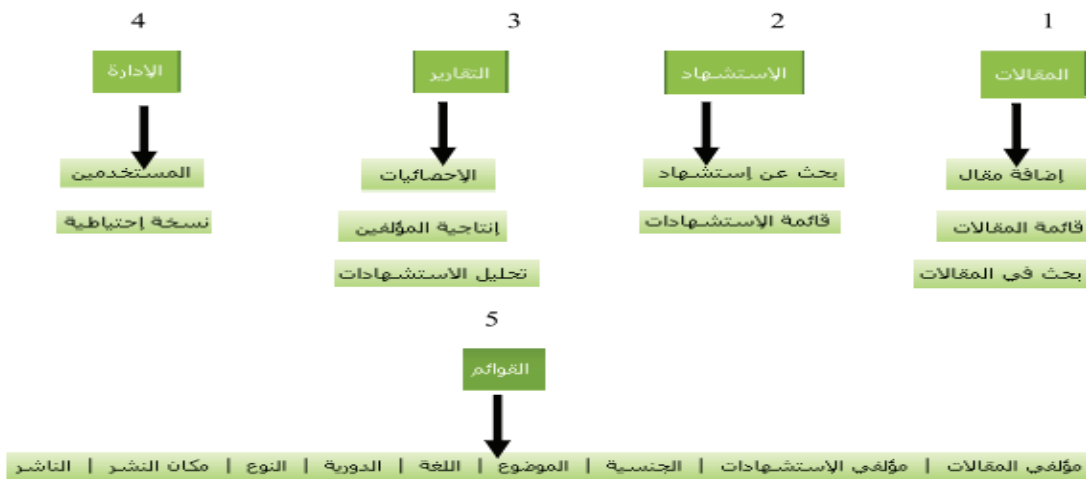


Fig. 1 The Five Main Elements and their Sub-fields.

1. *The main screen for outputs of the software:*
SAACI provides general statistics in the first screen for each user who has an account on this software. This part includes the total of number of studies that belong to the user, the

total number of articles which he/she has inserted, over the total number of citation.

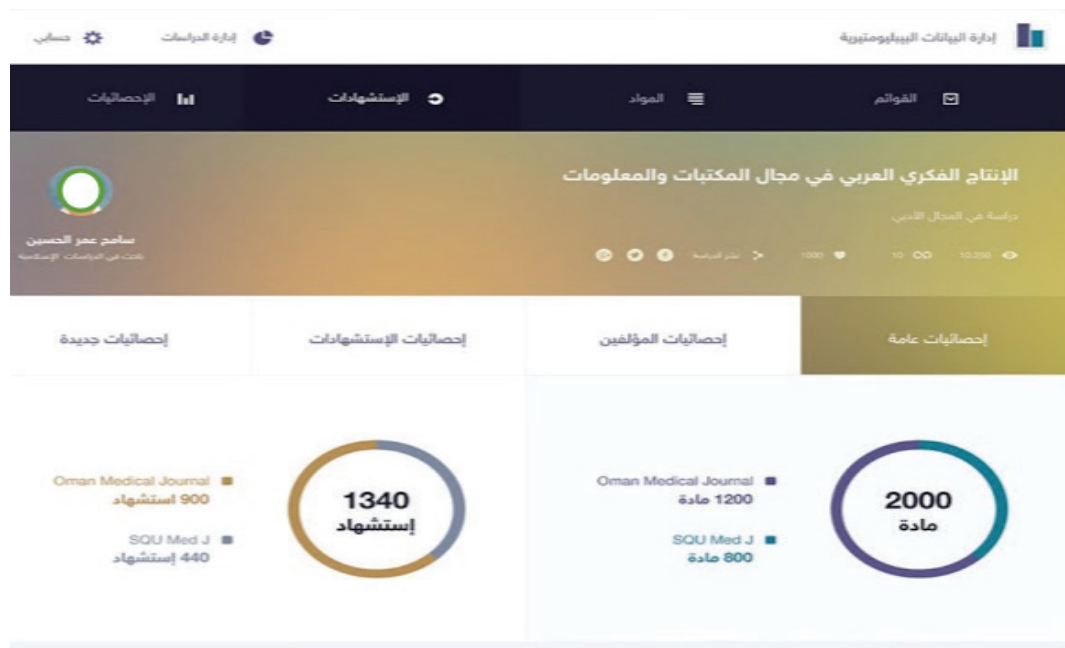


Fig. 2 General Statistics

2. *Authors characterized in articles:*
SAACI supports Arabic articles and its references and it helps the researcher to get the most important reports. SAACI divides reports into three parts: 1) Authors characterise in articles which include as follows:

- The most author published in the articles.
- The most author published in the references / citations.

- The most author published in articles / citations according to nationality.
- The rate of productivity of authors in the articles / citations
- Self-citation by authors.
- The rate of knowledge sharing among the authors in article / citations.



Fig. 3 The Characteristics of Authors in the Software

3. Authors / journals characterized in the citations:

This part in the SAACI provides the most popular reports based on bibliometric laws, which includes the following:

- The author who has the most citation in his articles.
- The author who has the most citation in his articles according to nationality.
- Self cited by journals

- The core journals
- Article distribution by subjects
- Article distribution by subjects in the citations
- Article distribution by publishing date
- Citation distribution by publishing date
- Citation distribution by sourcing types

The following figures show different reports which are available in SAACI.

التوزيع الزمني للمقالات

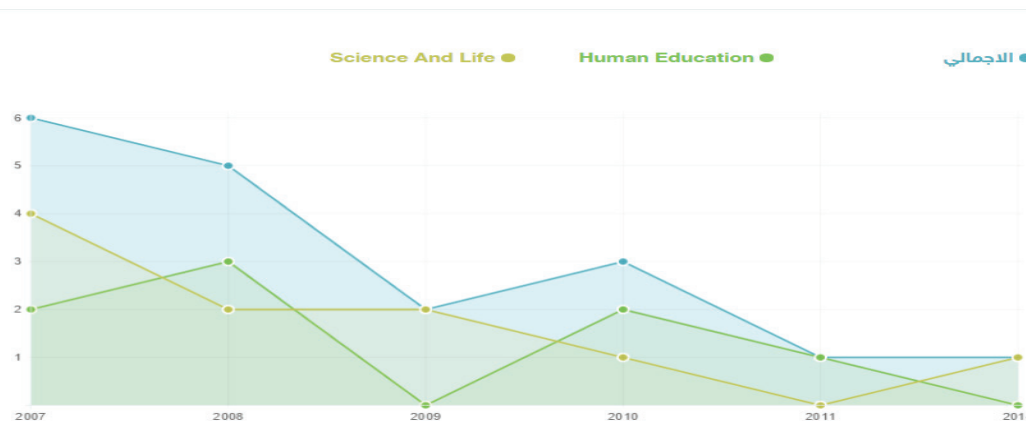


Fig. 4 Article Distribution by Publishing Date.

4. The fourth part of the reports provided by SAACI is the following:

- Bibliographic Coupling; a semantic similarity measure for documents that make use of citation relationships
- Cited Half-life

- Bradford Law of Scattering
- Obsolescence
- Impact Factor: the figure below shows the impact factors for Science and Life journal in 2009. This result can be based on the number of Arabic articles cited in it.

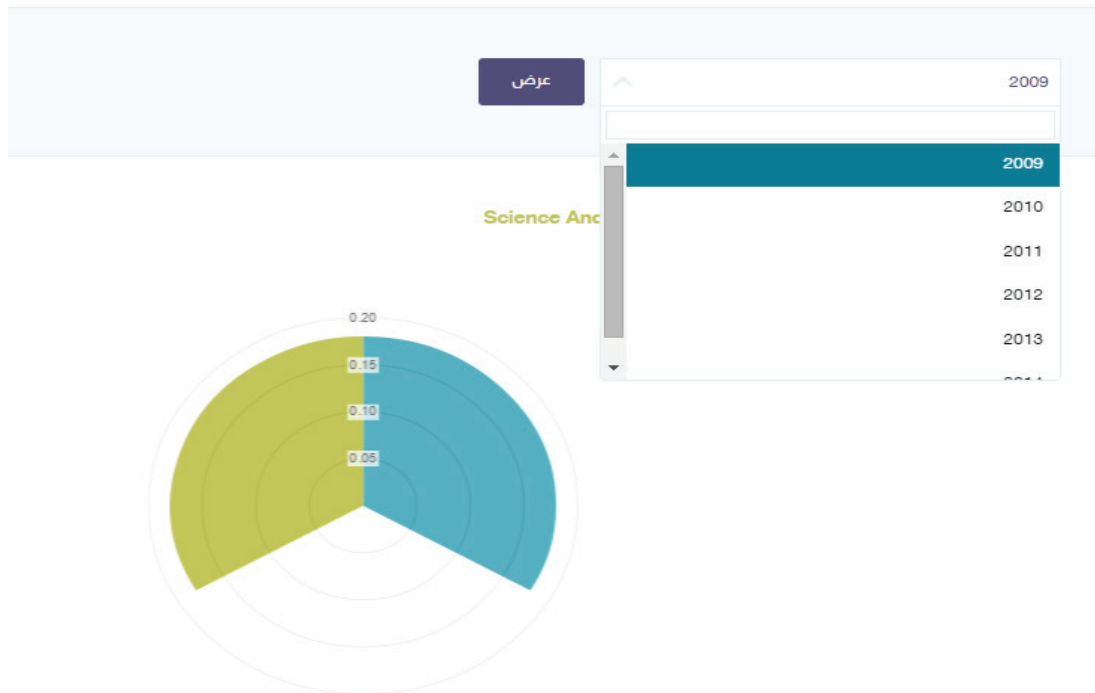


Fig 8 Impact Factor

XXXVI. CONCLUSION

This study identifies the issues and challenges in Arabic citation index. The results of the study indicate that the Systematic Analysis for Arabic Citation Index (SAACI) is the first Arabic open source software in the area of analyzing citation. The study demonstrates the advantages of the system which are easy to use and develop, due to it being an open source, and its suitability to all types of Arabic studies that focus on analysis of citation reference the program also analyzes most of the bibliometric measurements.

ACKNOWLEDGMENT

The authors wish to thank the Research Management Center and the Kulliyah of Information and Communication Technology in the International Islamic University Malaysia for their supporting.

REFERENCES

- [1] Qasim, Heshmat (1984). Analysis of reference citations and the evolution of paper measurements. Cairo: strange library.
 - [2] Timraz, Ahmed Ali (1991). Intellectual output is used by Arab researchers in the fields of engineering properties in Saudi Asaudih.alm books, 12, (2), 42 - 49.
 - [3] Al-Dosari, Fahd Mosfer (1991). Scientific contact at Arab researchers in pure science. Riyadh: King Fahd National Library
 - [4] Turkistani, Mohammed (1993). Intellectual production in the field of library and information: Bibliometrics study. Alam Al Kotob 14, (4) .462-463.
 - [5] Haddad, Faisal Abdullah (1993). Studies and reference citations in the magazine Alam Al Kotob: Bibliometric study. Alam Al Kotob 14, (4).
 - [6] Qandil, Yusuf (1994). Reference citations in the Journal of Resalat Almaktabah. Resalat Almaktabah , 29 (3) 23.
 - [7] Hamdi, Amal and Ghneim, Mohammed (2002). Electronic publication in ten years (1990-1999): Bibliometric study. Arabic studies in library and information science ,7, (2) 63-112.
 - [8] Khalifa, Abdul Sattar (2005). The use of an electronic source for information in the area of library and information through analysis of citation references with an internet sources in Arabic Periodicals articles. Arab 3000, 5, (3) 111 to 129.
 - [9] Zayed, Yusreya (2003). Electronic resources available in remote reference citations: analytical study of these libraries, documentation and information department ethics Cairo: 1998-2003.
 - [10] Alumer, Haifa Ali (2004). Study an intellectual production properties in the field of information technology through the analysis of the reference citations in Arabic periodicals. (Master) Faculty of Social Sciences: University of Imam Muhammad bin Saud Islamic University, Riyadh.
 - [11] Alzeheimi, Saleh (2007). Intellectual production in Medical Sciences in Oman properties (1996-2006) Bibliometric study, (Master Thesis submitted to the Department of Library and Information Science. Sultan Qaboos University).
 - [12] Kelwo, Sabah (2009). Automated systems applications in the field of bibliometric studies and their impact on the creative management of libraries. Available on http://www.slideshare.net/SalehAlzuhimi/savedfiles?s_title=ss-14697719&user_login=SabahKallow
 - [13] Najjar, Mohamed (2006). Programs mechanism for the formulation of reference citations. Tenth conference for specialist libraries in Egypt. Cairo.
- Alzeheimi, Saleh (2008). Bibliometric studies management system: a descriptive study. Conference of the Special Libraries Association branch of the Arab Gulf, Kuwait