

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Advantages and Disadvantages of the Webometrics Ranking System

*Suad Kunosić, Denis Čeke and Enver Zerem*

## Abstract

Today, there are several well-known global ranking lists for ranking universities in the world. While some of them ranked only a few hundreds of best and most influential universities, there are those that include a much larger number of ranking scientific institutions. One such global list which ranks the largest number of scientific institutions and scientists in the world is called Webometrics list. This list is very important for less developed economies and developing countries which have not established a sufficient quality control system of higher education quality, so this list serves as a corrective to the international evaluation of a wide range of universities in the world. In such a complex IT system of ranking an extremely large number of institutions and scientists, this system shows some disadvantages when ranking, which of course can be overcome by introducing certain improvements within the system of ranking. Systems that perform the collection, analysis, and indexing data have their advantages and disadvantages, which can sometimes lead to a misinterpretation of the data collected. Among other things, we will consider the possible solutions which would improve the rating system and prevent possible manipulation and uncertainty in the presentation of current and final results ranking.

**Keywords:** webometrics, university ranking systems, ranking scientists, university science transparency, web crawler, web scraping, Internet bot

## 1. Introduction

According to the definition by Björneborn and Ingwersen [1, 2], webometrics represents a joint (synergy) activity, i.e., application of other two approaches in one element known as bibliometrics and informetrics, for researching of the web, its information resources, structure, and technologies.

The name “webometrics” was defined in 1997, and it was created and established by Tomas Almind and Peter Ingwersen with an intention to show that informetric analysis can be applied to the web as an important source for measuring values (weight/sense) of documents and information [3]. They suggested several specific informetric parameters such as hyperlinks per website and sensitivity of links on websites distributed via types of documents and names of domains. Björneborn and Ingwersen [1] defined webometrics as “The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web, drawing on bibliometrics and informetrics approach.” One developed a detailed topology of links, a diagram of nodes on the web, and additional

terminology [4]. The area of activity and what webometrics includes may, in a wider sense, be characterized as (a) website content analysis, (b) web technology analysis, (c) web usage analysis, and web link structure analysis. Researches in this field imply creation of new discoveries based on analyses of numbers and types of hyperlinks, structure of the World Wide Web, and patterns of usage of the web as a mass communication medium and exchange of information.

Another definition of webometrics was given by *Mike Thelwall* in 2004, "... the study of web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study," which underlines development of applied methods for usage in a wider scope of social sciences. The purpose of this alternative definition was not to replace the primary definition within Information Science [5] but to support publishing of appropriate methods out of the scope of the Information Science.

After these events, Ingwersen represented Web Impact Factor (WIF) [6] in 1998, which represented a key metrics to measure and analyze hyperlinks of websites.

Basically, WIF measure may be defined as a number of sites of a certain web host (website) or portals referred to by links from other websites or web hosts, divided with the number of shared/published websites on that web host which are available to the web/web robots.

Namely, it was logical to assume that areas of great interest would attract more sequential links than average. The greatest advantage of WIF, in which logic was inherited from significance of quoting within an impact factor of a journal, was that it was easy to calculate with the application of advanced inquiries of a browser represented by AltaVista, a leading commercial search engine of that time.

However, usage of WIF measure was neglected after comprehensive analyses and obtained results, due to mathematical artifacts obtained from mathematical law, i.e., due to *power law distributions* of these variables. Other similar indicators which applied a size of an institution instead of a number of websites showed as much more useful for the purpose of analyses.

Subsequently, analyses of links were more directed toward analyses of influence of links and analyses of net connectedness of links, taking into consideration quantity of links as a reflection of productivity of researching within academic and scientific space.

Webometrics gradually evolved and became a great coherent field within the Information Science, at least from bibliometric perspective [7, 8], including analyses of links, web quotations, and a range of other web-based quantitative techniques.

Hyperlinks on websites are structured similarly like quotations in academic (scientific) journals since they guide from a source document to the final document. Similarity of links and quotations, together with the fact that universities were among the first ones which massively started applying advantages of the web, resulted in the appearance of numerous important naturally imposed research objectives. Such sequence of events imposed the question whether it was possible to use hyperlinks in a similar way as academic quotations or quotes in journals and articles, the question of validity of usage of a number of links and data obtained by AltaVista browser in the research and the best way to count links [9]. Simultaneously with these analyses of links, other researchers from the field of Information Science researched reliability and coverage of an area of browsing by browsers and change in contents of the web itself or individual collections of websites [10]. These three types of researches related to measuring the web are all together called webometrics.

Among other things, webometrics has become a useful methodology in many other fields, such as creation of ranking of universities in the world based on webometrics [11, 12], for scientometric evaluations or researches in some scientific fields [13].

## 2. Webometrics methodology

Webometrics methodology includes analyses of links, analyses of web quotations, evaluations of results of browsing via web browsers, as well as some basic descriptive studies and analyses of the web [15].

The web is of great importance as a communication medium, i.e., a platform for placing and archiving a wide spectrum of documents. A significant number of repositories of various kinds of documents are related to the academic society, and therefore application of this methodology in general ranking of a current situation in the academic field is more significant. Because of this huge and easily accessible source of information, unlimited possibilities for measuring or counting contents of the web turned up in a wide range (e.g., a number of web hosts or a number of websites) or in a narrower range (e.g., a number of web locations in a state, a number of web pages on a website of a university).

Although the terms “Internet” and “web” are usually treated as synonyms, they are not the same. Namely, the “Internet” represents a global network of computers which can share information, while the term “web” specifically refers to a group of interrelated documents available for review and downloading using HTTP [14].

For an analysis of the system of ranking of universities, Webometrics Ranking, the most important parts of webometrics methodology would be link analysis and web browser analysis, i.e., a survey of ways in which relevant information that are used in the ranking process of universities according to the Webometrics ranking methodology may be obtained.

### 2.1 Link analysis

Link analysis is a quantitative study of hyperlinks among websites [15]. Similar to a mechanism of counting of quoted works in journals and articles, the importance of websites can be evaluated with links and their analysis. As previously mentioned, the importance or influence of a website on the Internet is defined as Web Impact Factor [5], which, obviously, is similar to the concept of an impact factor of a journal (*Journal Impact Factor*, JIF).

The idea behind the link analysis is that one can practically assume that a number of links pointing to (guiding toward another location) the academic space of a web content can be proportional to productivity of a research of an organization at the level of a university [9], departments [16], research groups [17], or individual researches [18].

Calculation of WIF [19] can be obtained as a logic amount of a number of links, i.e., external or incoming links (*inlinks*) toward a website divided by a number of sites of a certain web host in a certain moment of time. Further information regarding the calculation of WIF is to be looked for in the work [6, 19].

### 2.2 Webometrics tools for collection of data from the Internet

Web tools such as search engines, web crawlers, and webometrics software which are used for collecting data from the web are called *Webometrics tools* [5].

The area of research in the field of webometrics can, in a wider sense, be divided into the following segments:

- Analysis of contents of websites
- Analysis of web technologies
- Analysis of application of web contents
- Analysis of structure of web links

To analyze data for the needs of webometrics, it is very important to know the source of information for each of the mentioned categories of webometrics. The main role of web browsers is to grasp relevant information on the basis of specific inquiries from various (heterogenous) sources of information.

Basically, there are two categories of sources of information which can be used in the research of webometrics:

- Commercial web search engines
- Personal web crawlers

Web search engines are computer programs which, on the basis of special algorithms, find appropriate information on the web, index them, and place them into databases appropriated for those purposes.

From the point of view of the webometrics research, web search engines can fundamentally be divided into two categories:

- Web search engines which support searches related to the field of webometrics
- Web search engines of a general type which do not have any additional capacity to direct searches toward terms related to the field of webometrics

Web search engines, such as Google, Yahoo, and Bing, enable users to access a vast quantity of information related to contents and structure of links on the web free of charge. Web browsers collect information in a similar manner as web crawlers which are used by users to collect linked data. Basically, web browsers contain three different parts, crawler, indexer, and interface, in which one enters inquiries with terms to be browsed [10]. Led by this fact, authors Aquillo et al. [11] applied advanced options of web browsers to collect data from the web for the needs of ranking of universities.

Web crawlers are programs with the main objective to collect data from precisely defined web locations. They function in the following way: they start collecting data from a certain web location, and then they apply links contained by that web location so that a web crawler could move automatically and independently to a next processed web location, from one site to another until there are several links to be monitored and analyzed.

Regardless of the existence of some additional tools for analysis of links such as *LinkDiscoverer* [20], *SocSciBot* [21], and *Webometric Analyst/LexiURL Searcher* [22], Thelwall and Sud [23] underlined that researchers still depended on *application programming interface* (API) of commercial browsers for collection of raw data for their webometric studies. These API functions enable automatic data collection and enable programmers to encrypt programs with which one can access results of browsing. Yahoo canceled its *free-of-charge* support for usage of API functions for the purpose of browsing, Google has limited access to its API from 2011, and Bing also has limited a *free-of-charge* access to API 2.0 from 2012. This essentially canceled or significantly limited possibilities to collect important information for



extensive researches within the field of webometrics. Although web browsers have a very important role in data collection, none of them is able to collect data from the whole web. The web is a dynamic environment, and there are fluctuations in the results obtained by browsing.

Generally speaking, one can say that web crawlers are an essentially better tool than web browsers if one talks about researches about the webometrics.

### *2.2.1 Data collection with commercial search engines*

The most popular web search engines, which are very popular besides their application in the webometrics, are Google, Yahoo, and Bing. Each of the web search engines uses its own algorithms for browsing and different techniques for indexing and browsing of the web. Actually, it means that if a user wants to enter an inquiry into a search engine in a form of, for example, “webometrics methodology,” there is a huge probability that he/she would obtain different results from different search engines for the same browsed term. These algorithms applied by the web search engines are business secrets of corporations standing behind their implementation. Besides the abovementioned search engines, there are other search engines, but these three are the most popular due to the quality of obtained results and speed of browsing. In application of some of web engines, there are some keywords for browsing that may be entered so that obtained results could be filtered and oriented toward a searched term. For example, if one enters the term “site:untz.ba” in Google search engine, the inquiry will provide us all data related to that domain, its auxiliary subdomains, and all sites indexed by the browser.

Furthermore, if one enters a string in the form of “site:untz.ba <space> filetype:pdf,” the browser would provide us all sites and subdomains containing documents of *Adobe Portable Document Format* (Adobe PDF) type and a direct link to the same. These examples are specifically applicable to Google search engine.

Web search engines of the Internet are very important in researching of the field of webometrics because their databases are a source of information that cover a great part of data of the web. Although commercial search engines are very important for surfing the Internet and data collection, they have some significant limitations, among which the following stand out:

- Web search engines do not index the whole web space [24, 10, 19].
- Systems for ranking of results of browsing eliminate similar or identical sites in results of browsing with an objective to eliminate useless information [25, 26].
- Algorithms that search engines use for surfing the web and generation of reports are corporative business secrets, and, therefore, an exact criterium for collection, sorting, and ranking information by importance is not known [19].
- A total result obtained in a search by a web search engine is assessed by time necessary for a search rather than by thoroughness and going into details into accurateness of data since they apply an algorithm which performs prioritization [4].
- Results may be conditioned by a national or a language area [27].
- Results may fluctuate and change from time to time.

Regardless of their limitations, commercial web search engines are one of the unique and best sources of information which are currently available but only for

certain types of webometrics researches. At the same time, they are not properly designed for the purpose necessary for the academic community, and results usually are not thorough enough, which would be a great need of this field [28].

If an interface of a web browser is directly used for browsing, data collection may be a very demanding process in regard to time. This problem may be overcome with an application of a special software based on application program interface developed by companies which create web search engines and other services on the web.

### 2.2.2 Web crawlers as a source of data

Another important source of data is personal *web crawlers*. Among the most popular free-of-charge tools of this type, which are used to analyze links, are *SocSciBot* [21] and *LexiURL* [29]. Both of these *crawlers* are developed by Professor Mike Thelwall from the University of Wolverhampton, UK, in order to find alternative strategies and methods to analyze links. The essence of functioning of these tools is that they search for and download certain websites from the web and analyze them with an integrated analytical software, such as *Pajek* [30], *Ucinet* [31], *NetDraw* [32], etc., for the purpose of data analysis and creation of a graph of a network representing a scheme of data linking.

### 2.2.3 Challenges within the webometrics research field

Webometrics functions on a principle of an analysis of academic and non-academic articles. Academic documents include publications such as e-journals, e-books, patents, technical reports, etc. Nonacademic documents include websites—commercial ones, sites of social networks, etc.—published by individuals, blogs, and portals where there is not any (i.e., which process of publishing of contents does not comply with) *peer-reviewed* system. The greatest challenges within the webometrics research field are in finding relevant sources of data and in the development and implementation of techniques for their efficient collection. Among the four research fields within webometrics, link analysis has been in the focus more and more since most of the commercial web search engines canceled their support related to browsing the web contents which include link analysis. For these reasons, there is still a great need for alternative sources of data.

## 2.3 Alternative sources of data

In the first decade of the twenty-first century, most of the web search engines supported the webometrics research filed with application of special keywords for search engines such as “site:domain,” “linkdomain,” “linkfromdomain,” etc. Starting from 2012 there has been a great change which has reflected the field of usage of sources of data for webometrics researches as a *matter of policy* which was started by owners of the web search engines. As a result of the mentioned, most of the web search engines canceled their support to the webometrics. Researchers from the field of webometrics tended to find alternative sources of data to go on with their researches. A survey of some of the existing systems by which data may be collected for the needs of webometrics analysis is given hereafter.

*Alexa Internet* [33] was established in 1996. As a *search engine optimization* (SEO) tool, Alexa collects data on the basis of behavior of users on the web, while they visit some sites using their analytic tool. The data are analyzed in a manner to give information for a global ranking or ranking within a country. One also analyzes data related to web communication and a total number of sites which refer to a certain web domain, etc.

*Alexa Toolbar Service* [34] is a smaller software application which collects and stores information about websites, web domains, and other sites which this tool uses to collect data regarding analyses of users.

In 2005 *Who.is* [35] become a web portal for searching for and collection of data about web domains of any organization or institution. *Who.is* offered a unique tool to obtain information about IP addresses, locations of domains, DNS names of servers, information related to availability of domains, and information related to various organizations or universities which belonged or belong to those institutions.

*Webconfs.com* [36] represents one of the additional tools which may be used as a source of data for webometrics researches.

*Majestic SEO* tool [37] represents one of the best tools related to analyses of *back-link*, *incoming link*, *inbound link*, *inlink*, and *inward link*. Backlink for an assessed web resource is a link which shows a hyperlink from some other web location to the observed web location. A web resource may be a web host, a website, or a web directory. Backlinks are one of the indicators of popularity of a website, and they represent a very significant source of information. A rank or value of a site within a web domain increases depending on the quality of *backlinks*.

*Searchmetrics* [38] is a professional SEO tool which enables a survey of all data related to visibility and social visibility of websites. Visibility of a site is analyzed through *PageRank* [39], which is a tool for analysis of *metatags*. Afterward one analyzes a server and a domain where a certain content is located (domain's age, domain's popularity, reverse IP addresses), if there are tools for analyses of links (popularity of a link, counter of backlinks, value of links, exit links). Visibility of social data is related to links related to social networks such as *Facebook*, *Twitter*, *LinkedIn*, and *Google+*.

*Ahrefs.com* [40] is a famous set of tools (*Site Explorer* [41], *Content Explorer* [42], *Keywords Explorer* [43], *Rank Tracker* [44], *Site Audit* [45]) for analyzing backlinks to the websites, and it is a very important tool for SEO analyses.

### 3. University ranking systems

These days, the Internet has become the main source of scientific information, both for the academic community and for the society. The whole society has been turning to the Internet as a primary medium for presentation of information to the public. On that ground, the fact that web publications are a primary tool for communication within the educational system and that they reflect the complete picture of quality and performances of universities has become very important [51]. Bearing in mind the development of digital world, the influence of electronic publications is significantly greater than the influence of written media or printed versions of journals and books today. Websites are the cheapest and the most efficient way to stimulate all of the three academic missions: to educate, to research, and to transfer knowledge [51]. This fact is one of the main reasons why web data have been extensively used for evaluation, inter alia, of universities and research institutions in the last couple of years.

Ranking is a process in which one defines positions of elements in a group in regard to a total system so that for any two elements in a sequence, the first one is ranked "as higher than," "as lower than," or "as equal to" the second item of a sequence [46].

Ranking process appears in many fields whether they are academic or of other type. In a case of academic space, ranking may be applied in different parts of academic space, starting from ranking of professors and ranking of researchers and research centers to ranking of universities. Ranking of universities is an especially interesting field of application of ranking.



Currently, one implements a process of comparing and evaluating universities in the domain of academic and research performances with the existing system of ranking of universities. Most of the academic institutions rely on data obtained from the ranking system of universities which serve them as indicators of a progress of an institution over time in regard to other academic institutions [47]. Besides this, information from these ranking lists often serve as a basis for applying for and obtaining financial assets from founders or other institutions on the basis of a position on these lists [48]. On the other hand, potential beneficiaries of services of a university use these lists to evaluate academic institutions to decide which one to attend and to evaluate which one provides better options for education and further employment.

The study [47] identified 24 ranking systems. Thirteen ranking systems were analyzed into details since their lists were active during the last couple of years, i.e., from 2015 to 2016. Other ranking systems were excluded from further analysis because they did not publish information and did not include indicators of their performances or published their ranking methodologies. The study evaluated between 500 and 5000 institutions. The oldest ranking system, *Carnegie Classification*, was established in 1973. All other ranking systems were first published between 2003 and 2015. The study mentioned that three ranking systems were led by universities, two were led by agencies, five by consultancy or independent groups, and one was led by an institution established by a government.

In the analysis from the study [47], 4 systems for ranking out of 13 for evaluation claimed that they used their results to evaluate quality or performances of researches. Nine of thirteen systems use a total number of publications as an indicator for evaluation of quality or performances of researches—this is usually defined as a number of *peer-reviewed* articles from bases of Thomson Reuters' Web of Science Core Collections or SCOPUS which is maintained by Elsevier. On average, 33.8% ranking results are ascribed to publications and quotations or to various versions of these metrics [47].

Ranking systems that strongly rely on metrics related to publications and quotations are *Leiden Ranking*, *Shanghai*, *SCImago*, *URAP*, *US News and World Report*, and *EU U-Multirank* systems. The fact that SCImago ranking system takes the presence on the web into consideration by Google metrics [49], which is 20% of the total result of ranking, is very interesting. Similarly, Webometrics ranking list includes all universities of the world which are present on the web in the ranking system. The objective of this list is to encourage universities and their personnel to increase visibility of universities through creating more websites of university organizations and institutions. A survey of percentual participation and application of individual indicators applied by various systems is given in the work into details [47]. According to [47] current indicators are not adequate for an accurate assessment of results of researches, and they need to be amended and expanded to satisfy a standardized criterium.

### 3.1 Webometrics Ranking of World Universities

Several research teams have been working on the development of web indicators since the mid-1990s. Realizing possibilities of this kind of ranking, the European Commission started several projects for this purpose: EICSTES ([www.eicstes.org](http://www.eicstes.org)—currently inactive), WISER ([www.wiserweb.org](http://www.wiserweb.org)—currently inactive), and [www.webindicators.org](http://www.webindicators.org) (currently inactive).

After noting capabilities and importance of web search engines as the main agent to access information being processed and being processed by the web [53], one created new indicators [54, 11] which should have been milestones to solve

problems arising from instability of results of browsing via web search engines [55] and artifacts arising from calculation of Web Impact Factors [6] [51].

The first catalogs of universities were created with projects EICSTES and WEISER, and the first preliminary list of these universities based on web indicators was published in 2004. This application of cybermetrics or webometrics techniques did not significantly differ from similar scientometric proposals where bibliometric data were the basis of information used in analyses [56, 51].

Most of the bibliometric indicators, such as a number of publications or quotations, are easily available. However, the problem with such access is that in this way only a limited number of information about activities, researchers, and observed institutions are available since only formal publications are taken into consideration. Actually, scientometric tasks should contain more elements, and more variables should be added [51].

However, including additional elements in an analysis, particularly when they are not easily available, may complicate the analysis and sometimes may be inapplicable when it comes to a global work plan. Among other things, there is an attitude that publications are not the only indicator of evaluation of professors. There are, inter alia, materials for lectures, raw data, slides from lectures, software, and bibliographic or linked lists (bookmarks), which are also deemed as relevant information about a professor's dedication to students [51].

Besides these data, a structure and a content of all kinds of administrative information provided by an institution also have their value. All these elements speak for themselves when published publicly in the virtual world, the web world, and are very good indicators of an academic level of an educational institution. The fact that if someone is not on the web she/he does not exist supports the previous statement. Web space provides a comprehensive way to describe a wide range of activities of an institution where scientific publications represent only one of components which may be found on websites.

Today, highly ranked researchers, institutions, and universities publish millions of pages with various materials composed of hundreds of departments and services, hundreds of research teams, and thousands of students on their websites.

Until now one has talked about webometrics methodology and systems for ranking of universities generally. However, the topic of this chapter is oriented toward a specific system of ranking of universities which applies webometrics methodology for the world's ranking of universities. This chapter will elaborate on Webometrics Ranking of World Universities, which was developed by and is under the competence of Cybermetrics Lab (Spanish National Research Council, CSIC) [50], who developed indicators called web ranking (WR) for the ranking process and who initially considered the following elements in the ranking process [51]: a number of published websites (S); a number of files contained, including PDF, ps, doc, and PPT form of documents (R); a number of articles collected via Google Scholar (GS) database system (Sc); and a total number of external links (V).

Webometrics Ranking of World Universities is the largest list for academic ranking of higher education institutions. From 2004, Cybermetrics Lab has implemented an independent, objective, free, open scientific exercise for provision of reliable, multidimensional, updated, and useful information about performances of universities from all over the world on the basis of their presence and impact on the web every 6 months.

*Cybermetrics Lab* has been developing quantitative studies on the Academic Web Network since the mid-1990s. The first indicator was introduced during the EASST/4S conference in Bielefeld (1996), and collection of web data from European universities started in 1999 with a support of EICSTES project financed by the European Union.

These efforts are a continuation of scientometric research Cybermetrics Lab which started in 1994 and which was presented on a conference of the International Society for Scientometrics and Informetrics (ISSI, 1995–2011) and International Conferences on Scientific and Technology Indicators (STI-ENID, 1996–2012) and published in journals with a great impact effect (*Journal of Informetrics*, *Journal of American Society for Information Science and Technology*, *Scientometrics*, *Journal of Information Science*, *Processing Information and Management*, *Research Assessment*, and others). In 1997 one started issuing journal *Cybermetrics* dedicated to published works about webometrics.

After publishing of ranking of the University of Jiao Tong in Shanghai, Academic Ranking of World Universities (ARWU) [52] in 2003, team Cybermetrics Lab decided to approve the main innovations proposed by Liu and his team. It was suggested that ranking should be done on a basis of publicly available web data, combining variables in a composite directory and with a real global coverage. The first edition was published in 2004, and it has been issued two times a year since 2006. After 2008 the portal has included webometrics ranking for research centers, hospitals, repositories, and business schools.

### 3.1.1 Composite indicator

Probably one of the most important contributions of *Shanghai* ranking was introduction of the **composite indicator**, which combines a system of weighing factors with a set of indicators. Traditional bibliometric indexes are made on ratios such as Garfield's journal impact factor which is based on variables which follow the power law and are useless for description of huge and complicated scenarios.

Ingwersen's proposal from 1998 [6] for a similarly designed *Web Impact Factor* which uses ratio *links/websites* ( $L/W$ ) was equally useless due to mathematical artifacts which it generates.

Following the *Shanghai* model up, Cybermetrics Lab developed an indicator which transforms relation  $L/W$  into the following formula  $aL + bW$ , where  $L$  and  $W$  should be normalized in advance and  $a$  and  $b$  are weights which add 100%. Cybermetrics Lab strongly discouraged the usage of WIF due to its serious disadvantages. The composite indicator may be designed with different groups of variables and weights according to the needs of programmers and models. Webometrics applies "a priori" scientific model for the creation of a composite indicator. Other ranking lists chose arbitrary weights for very dependable variables and even combine raw values *with ratios*. None of them follows up a logic relation among variables related to activities and influential variables, i.e., each group represents 50% out of the total measure of weight.

Values should be normalized before any combination of variables, but the practice of application of percentage is mainly inaccurate due to *power law* distribution of data.

Webometrics *log* normalizes variables before combining in the ratio of 1:1 between activity/presence and visibility/influence of a group of indicators.

### 3.1.2 Collection of data for webometrics ranking

Collection of a great quantity of data from the Internet, where one has to go through thousands of sites, may be done only automatically. One of the possibilities is to use commercial or free-of-charge crawlers, but adjustment of such systems for adjusted needs may be a complicated and difficult task, and it requires a significant participation of human and computer resources [57]. On the other hand, web search engines already have well-designed and tested systems for this need,

and they do regular updates of their databases and have many tools which enable automatization of work so that machines may be easily adjusted to extract required data. Furthermore, web search engines are the main agents in navigation process on the web, and therefore the presence of a web domain in their databases represents an indicator of visibility on the Internet. Commercial web search engines also have limitations, which often include inconsistent and rounded-off results of browsing, favoritism in geographic and language coverage of results of browsing, or frequent and nontransparent changes in their work procedures. Due to the mentioned problems, one uses several web search engines together in practice, when collecting data. The most popular search engines such as Google (and Google Scholar), Yahoo Search, Bing, Exalead, and Alexa [11] are used for these purposes.

3.1.3 The webometrics ranking weighing model

Webometrics ranking system [58] performs an evaluation and ranking of universities of the world two times a year (January/February and June/July) by its own developed methodology. Webometrics ranking methodology includes several phases and applies several systems so that data necessary for ranking and analyses may be updated and collected in time.

According to [51], there are three key aspects that need to be measured in the academic web space:

- Size, i.e., quantity of published information
- Visibility, number of certain cases of appearance on other web hosts which refer to the analyzed web host (quotations of websites-hosts = number of external incoming links) obtained by a domain
- Popularity, which represents a number of visits to a website

Bibliometrics has traditionally ignored frequency of appearance of a journal on various locations or sources of data and has focused on an impact of a journal, i.e., relation between a number of quotations and a number of published articles in the journal. A similar approach was proposed in the case of Webometrics ranking.

Webometrics ranking performs monitoring of a certain group of parameters (criteria) (Table 1), but only size and visibility of a web host are included in the

Criteria	Indicator	Sources	Weight
Size	Number of pages (S)	Google, Yahoo, Live, Exalead	25%
	Number of rich files (PDF, PPT, DOC, and PS) (R)	Google	12.5%
	Number of papers (Sc)	Google Scholar	12.5%
Visibility	Number of external links (V)	Yahoo, Exalead, Live	50%
Luminosity	Number of external outlinks		
Subdomains	Number of subdomains		
Popularity	Number of visits		

Table 1.  
Criteria and weights used in the calculation of the WR indicator [51].



final data which are used for ranking. A model for ranking defines that a relation between these two parameters (size and visibility) is taken in the ratio 1:1. In order to take diversity of academic activities and services into consideration, component “size” is divided into three parts so that one could measure raw data about a quantity of websites, a number of rich files, and a number of articles and publications collected by Google Scholar system.

According to the work [51], criteria and weights used for calculation of WR indicators in those times were obtained from several sources only, which mainly were web search engines. Some of those search engines are not used to obtain data any more, but there are some new search engines together with some of the old ones which improved their algorithms for indexing and browsing of results from the web.

Pursuant to the proposed model, ranking (web ranking) is calculated with the following equation (Eq. (1)):

$$WR = 2 \times Rank(S) + 1 \times Rank(R) + 1 \times Rank(Sc) + 4 \times Rank(V) \tag{1}$$

The ratio which combines weights ascribed to each of the elements is (2 + 1 + 1):4 or 1:1, which was the initial intention. In order to avoid problems related to size, search engine bias, and other factors, results collected in this way, which initially were expressed as absolute values of numbers, are log-normalized and transformed into ordinal numbers and then combined with the previously mentioned equation for WR [59].

Over the years of application of the system, Cybermetrics Lab has made adjustments of indicators of calculation according to the analyses of data available during the years preceding the analyses. The data shown in tables on [www.webometrics.info](http://www.webometrics.info) are basically ranks (smaller number, better ranking) which purpose is to show individual performances, but one has to bear in mind that those values are not applied in ranking calculations [60].

Due to technical problems in the previous versions of the ranking system, Cybermetrics Lab changed some of the ranking weights (presence and excellence from **Table 2**) in the last version of the ranking system so that the current methodology is shown in **Table 2** (January edition, 2019.1.0.).

Indicators	Description	Source	Weight
Presence	Size (number of web pages) of the main web domain of the institution. It includes all the subdomains sharing the same (central or main) web domain and all the file types including rich files like PDF documents	Google	5%
Visibility (or impact)	Number of external networks (subnets) originating backlinks to the institution's webpages After normalization, the average value between the two sources is selected	Ahrefs, Majestic	50%
Transparency (or openness)	Number of citations from top authors according to the source	Google Scholar Citations	10%
Excellence (or scholar)	Number of papers among the top 10% most cited in 26 disciplines Data for the 5-year period (2012–2016)	SCImago	35%

**Table 2.**  
*Webometrics university ranking methodology (January edition, 2019.1.0.) [60].*

### *3.1.4 Several relevant facts about webometrics ranking of universities*

Results of ranking of universities [58] have been published two times a year since 2004 (data are collected during the first week of January and July to be prepared and published in the end of both of the months) covering more than 28,000 institutions of higher education all over the world with their analyses.

The data are collected between January 1 and January 20, depending on a current edition of a ranking publication. Data are taken (sampled) for each of the variables at least two times during the certain period, and the greatest value is taken as the final value to be analyzed in order to avoid possible errors in data collection. Inconsistency of web search engines is very huge so that the obtained results may be diversified, and there is a small possibility for their replication if browsing is done several days later. Google is very geographically biased; that is why data are collected with *google.com* mirror related to the domain, and English is used for the interface and Madrid (Spain) as the location.

A final publishing of ranking data is performed at the end of January or July, usually not before the 28th day of the month. It is very important to mention that Cybermetrics Lab follows its general rule not to discuss any presented result or provide unprocessed data with which a specific ranking was performed [58].

## **4. Webometrics ranking system: advantages and disadvantages**

Like other ranking systems, Webometrics ranking system has a range of advantages and disadvantages. Differently from other systems of ranking of scientists and universities, one can say that webometrics is a “global” ranking system. Why global? Most of the ranking systems include only several hundreds or thousands of the best universities, such as Shanghai list, while Webometrics includes most of the universities of the world, i.e., currently 28,000 scientific institutions from all over the world [60]. This list also enables ranking of scientific institutions, institutes, and individual members of a university, which can entice competitive spirit among individual members of a university. Why is this important? An extremely small number of universities of the world satisfy the Shanghai list criterium. However, this does not mean that there are no other universities of good quality besides those which are ranked as well as scientists working at those universities all over the world. It is easy to conclude that the universities from the Shanghai list and similar lists mainly originate from countries from well-developed economics and well-ordered educational systems, developed democracies, and high degrees of autonomies of their universities. Higher education systems of developed economies follow up the needs of the labor market and technology progress, and the quality of educational institutions is institutionally maintained due to strict accreditation criteria prescribed by authorized organs and ministries in every state. In developing countries and in poorly developed economies, there are great problems regarding an objective assessment and ranking of quality of institutions of higher education due to:

- Poor or no applying of international criteria [61, 62]
- Involvement of politics into institutions of higher education [61, 62]
- Devaluation of diplomas and criteria through institutions which do not fulfill requirements and criteria prescribed by law [61, 62]
- An extremely low percentage of scientific production of relevant publications indexed in the leading databases

In such circumstances, Webometrics ranking system actually represents a system of neutral international evaluation of quality of scientific institutions and scientists at all institutions not included into the Shanghai list. It is important to underline that there are not any significant deviations in the placing of the first 100 ranked universities on the Shanghai list and on the Webometrics list. Namely, the Webometrics list, through the four of its criteria (**Table 2**) [60], evaluates situations at the universities all over the world and positions them on its lists assessing every of the four criteria individually. This ranking procedure cannot be affected by any university, ministry, or state trying to improve its institution's ranking position. In most of the low-developed or non-developed countries, there is not any adequate system to control and follow up the success of reforms or define weaknesses and evaluate destructiveness of the involvement of policy into activities of institutions of higher education. This is one of the great advantages of Webometrics because it actually represents a very simple international tool for quality control of higher education institutions and enables competition among higher education institutions all over the world. In such process, it is clearly visible through the ranking system which of the four ranking parameters (presence, visibility, transparency, excellence) an institution progresses or stagnates. This enables development of a strategy for improvement of quality of scientific institutions, particularly in the weakest segments being evaluated. An extremely good point of Webometrics is that it performs ranking of institutions within states, within regions, or within the whole world [63]. In total, a university can be better positioned in some country, but it does not simultaneously mean that it is better than others by all of the four ranking parameters.

According to the last Webometrics list for Bosnia and Herzegovina [64] (January 2019, Edition 2019.1.2.), the University of Banja Luka is positioned second in Bosnia and Herzegovina, but in "presence rank" category, the International University of Sarajevo (which, in total, takes only the fourth position in Bosnia and Herzegovina) is ranked better than the University of Banja Luka (**Figure 1**).

How important having an international ranking list as a corrective showing and assessing situation at higher education institutions in developing countries may be analyzed on the example of universities of Bosnia and Herzegovina. Bosnia and Herzegovina is composed of two entities (the Federation of Bosnia and Herzegovina and the Republic of Srpska), and the Federation of Bosnia and Herzegovina is administratively divided into ten cantons. At the level of the state, there is the Framework Law on Higher Education in Bosnia and Herzegovina, while at the level of entities and cantons, there are educational policies being implemented according to the laws of entities and cantons. There are 8 public and 35 private universities and faculties in Bosnia and Herzegovina which are ranked by the Webometrics list from January 2019 (January 2019, Edition 2019.1.2.) [64], which is an extremely great number for the country with about 3.5 million of inhabitants. Work permits and work control of these higher education institutions, without clear international criteria, are issued and implemented by cantonal ministries of education with laws differing from canton to canton. Cantonal laws often are not in compliance with the Framework Law on Higher Education in Bosnia and Herzegovina ("Official Gazette of Bosnia and Herzegovina" no. 59/07 and 59/09, hereinafter: the Framework Law), and very often they are subjects of dispute before the Constitutional Court of Bosnia and Herzegovina (rating of constitutionality: U-19/16, U-22/18) [65, 66]. Cantonal laws are often changed for the purpose of involvement of politics into the universities in order to weaken and cancel their autonomy which is guaranteed by the Framework Law on Higher Education in Bosnia and Herzegovina. In such conditions, the only measure and objective evaluation is the Webometrics list. Namely, the University of Tuzla has been progressing on the Webometrics list over the years

## Bosnia and Herzegovina

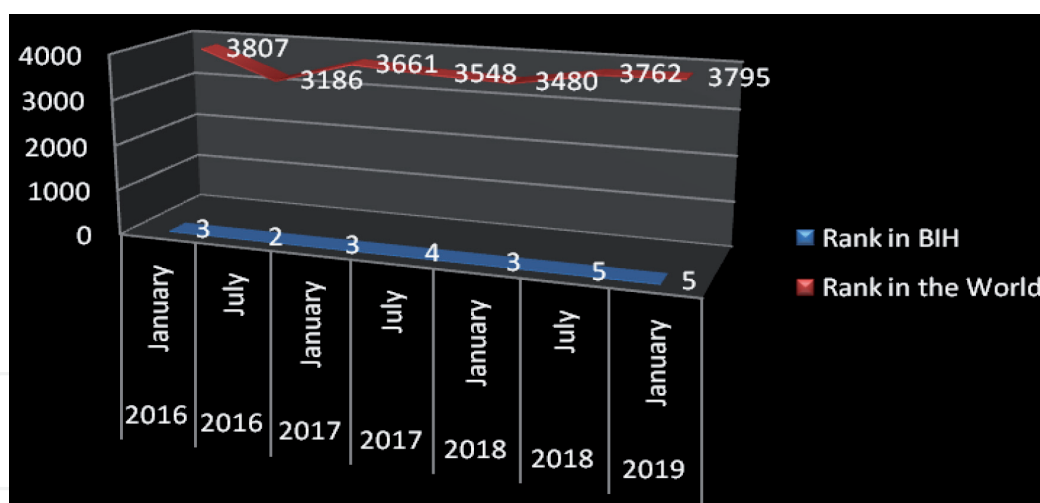
ranking	World Rank ▲	University	Det.	Presence Rank*	Impact Rank*	Openness Rank*	Excellence Rank*
1	1785	<a href="#">University of Sarajevo / Univerzitet u Sarajevu</a>	👉	560	2231	1419	2334
2	3416	<a href="#">University of Banja Luka / Univerzitet u Banjoj Luci / Универзитет у Бањој Луци</a>	👉	1198	7284	3848	3414
3	3619	<a href="#">International BURCH University Sarajevo</a>	👉	741	6427	4049	3925
4	3671	<a href="#">International University of Sarajevo</a>	👉	660	7245	4129	3816
5	3795	<a href="#">University of Tuzla / Univerzitet u Tuzli</a>	👉	2347	7938	3610	3816
6	4840	<a href="#">University of Zenica / Univerzitet u Zenici</a>	👉	2631	8413	6451	4666
7	5027	<a href="#">University of Mostar / Sveučilište u Mostaru</a>	👉	1302	13403	4257	4269
8	5694	<a href="#">Sarajevo School of Science &amp; Technology</a>	👉	17771	13223	3826	4566
9	6041	<a href="#">European University Brčko / Evropski Univerzitet</a>	👉	11223	4188	7607	6033
10	6635	<a href="#">Evropski Univerzitet Kallos Tuzla</a>	👉	12943	4225	9676	6033
11	6915	<a href="#">University of East Sarajevo / Univerzitet u Istocnom Sarajevu / Универзитет у Источном Сарајеву</a>	👉	1396	20114	5109	4269
12	9830	<a href="#">International University of Travnik / Internacionalni Univerzitet Travnik</a>	👉	4995	7225	11401	6033
13	11257	<a href="#">University Džemal Bijedic Mostar / Univerzitet Džemal Bijedic Mostar</a>	👉	4671	11854	9938	6033

**Figure 1.**  
The Webometrics list for Bosnia and Herzegovina with all parameters (January 2019, Edition 2019.1.2.) [64].

(after implementation of a set of measures by a quality team), and in June 2016, it took the 3186th position on the list, which was the best position of this university in the history (**Figure 2**). Immediately after this event, the Law on Higher Education in Tuzla Canton was changed, and in a day the management of the university was replaced, the Senate was dissolved, and receivership with temporary organs under political patronage was imposed. Although it is difficult to find such case in developed economies, or anywhere in the world, the key issue is how to measure the effect of such measure.

In circumstances where there is not any adequate reaction of state institutions (Agency for Higher Education, parliamentary committees for education, state Ministry for Civil Affairs) to such situation because everything is politically controlled and the Constitutional Court is declared authorized for interpretation of compliance of the Cantonal Law on Higher Education with the Framework Law on Higher Education in Bosnia and Herzegovina (U-19/16) [65], an independent international factor unaffected by politics, i.e., Webometrics, is necessary. Although such measure was allegedly implemented to improve quality and position of the University of Tuzla, the Webometrics list soon showed all the effects of this measure. The position of the University of Tuzla on the Webometrics list was becoming weaker and weaker over the years, and in January 2019, it ended on the 3795th [64] place (**Figure 2**) and experienced the fall for 609 positions or 19.11%. The university which in July 2016 (**Figure 2**) took the second position in Bosnia and Herzegovina by quality took the 5th place. One more advantage of Webometrics is ranking of four segments, which provides an insight into segments where the university became weaker and into those where it became stronger. The mentioned indicators show fall of quality in almost all of the ranked segments, and it imposes the conclusion that cancelation of autonomy, involvement of politics into the university, and compulsory administration cause weakening of the quality of the university. Similar processes and measures were implemented at the University of Bihać, which





**Figure 2.**

*Position of the University of Tuzla in the world and in Bosnia and Herzegovina according to the Webometrics list for the period from January 2016 to January 2019.*

resulted in its taking 11,546th position according to the list from January 2019 [64], and it would be very hard to improve its position significantly. This resulted in a significant decrease of the number of enrolled students, in the decrease of competitiveness, and in the struggle for huge international projects and low percentage of scientific production in the leading index bases of the world. An obvious decline of publishing and quoting of works is confirmed by the criterium of excellence by which the University of Tuzla takes the 4th place in Bosnia and Herzegovina [64], with a condition that the mentioned parameters probably are not correct which is a consequence of Google Scholar and webometrics itself. Although it measures the researching productivity on the basis of the presence on the web, the mentioned ranking system depends on categories related to publishing and quoting of scientific works (of excellence (SCImago) and transparency (Google Scholar)) [60]. These two parameters bring 45% out of the total score in ranking of a university. Over the years there has been a notable tendency of increase of the weight of these two parameters in Webometrics ranking system. Here we see first the disadvantages of Webometrics ranking system. Webometrics uses GS for transparency criterium, which enables creation of a profile of a scientist with verification of an address from a scientific institution.

Google Scholar is a system in which the academic community has been very interested [67], and it has been used by a great number of universities and research institutions both for ranking of institutions and ranking of academic personnel [68–72]. The system is very good; it is automatized in the way that a computer program performs the main role in the whole process, from data collection to data processing. Like any other system, this one is not perfect, and it has some critical omissions which are mostly related to ascribing a quotation from a scientific paper to some authors to whom it does not belong.

GS system has several possibilities to ascribe articles to their authors. The first possibility is automatic. Computer system collects information about published scientific publications on the web, with all auxiliary elements of the paper which besides the title of the work include names of the authors, keywords, and a brief description of the paper. According to these data, GS system browses its basis of user profiles and proposes to a potential author found in its database of publications, which contains a name of a potential author, to ascribe the found article to his/her profile. On the basis of available data, GS system “assumes” that the user is the author of the publication found, which does not have to be true. This often

causes situations where papers and quotes on certain profiles are not true. Of course, an author has to have his/her profile on GS system.

Another possibility is manual addition of a publication with all of its auxiliary meta-information. By this approach, GS system enables a user who has a profile on the system to enter data about his/her paper manually: a title of the paper, a list of all authors, a name of a publisher, a title of a journal or a conference, a year of publication, etc. There is not any mechanism to check authenticity of an author, i.e., if a person is the author or a co-author of the paper. The only good point of this approach, bearing in mind manipulation of quotes in the concerned publication, is that all quotes related to the paper concerned would not unconditionally be ascribed to the user of a profile, i.e., to the “author” who adds the paper to his/her profile. A possible explanation for this situation may be in the fact that there is not any possibility to acquire all quotes of a paper at once since it is done automatically by web crawler computers which have their time scheme of performance of tasks.

The third possibility is to add publications manually but not data. One browses the database of GS system and finds a desired publication and ascribes it to the user profile. This differs from the previous approach by the fact that the work concerned was already indexed in GS system and all necessary data (which besides the main data contain data about all quotes related to the publication) were ascribed to it. In this case, if a publication having a certain number of quotes has already been ascribed to an author who is its real author and if a user manually adds the publication to his/her profile, then all existing quotes of the publication concerned would be ascribed to his/her profile regardless of the fact that he/she is not the real author or co-author of the publication. According to our best knowledge, currently there is not any mechanism to check credibility of an author—if a person really is the author of the co-author of the concerned publication. This is one of the great disadvantages of the current version of GS system regarding manual ascription of publications to users' profiles.

Since we believe that the last mentioned situation represents a serious omission in GS system, we tested the concerned situation with two articles. One publication which we added to a profile of a user (with No = 47 quotes) who is not the author of the publication (one of the authors of this article) had a significant number of quotes (No = 388), while the other one did not (No = 20). Two publications with different numbers of quotes were added in order to check if it was really practically possible to add a publication with any number of quotes. We made a screenshot before adding the concerned publications to the profile of a user who was not the author of any of the articles (**Figure 3**) and after the publications were added (**Figure 4**). These two illustrations clearly show that after the concerned publications were added, the number of quotes of “new author” increased significantly proportionally to the number of quotes ascribed to the source publication. GS system did not, at any time, report that the “new author” actually was not the author of the concerned publications.

This represents significant omission in Google Scholar system, which opens possibilities for new ways of manipulation in all systems of ranking, of universities, and of researchers themselves, which use this system as a part of some other systems for various types of ranking. Some authors had already been pointing to manipulations with quotes in academic researches [73], but this is the case of manipulation of GS system [74, 75].

Where does this omission become very “disputable?” Namely, examining profiles of scientists from the University of Tuzla, one defined that even 30% of profiles out of the first ten ranked contain papers of which they are not authors or co-authors. Those are profiles taken into consideration in ranking of parameters on the list from January 2019. It is significant to mention that even 446 quotes

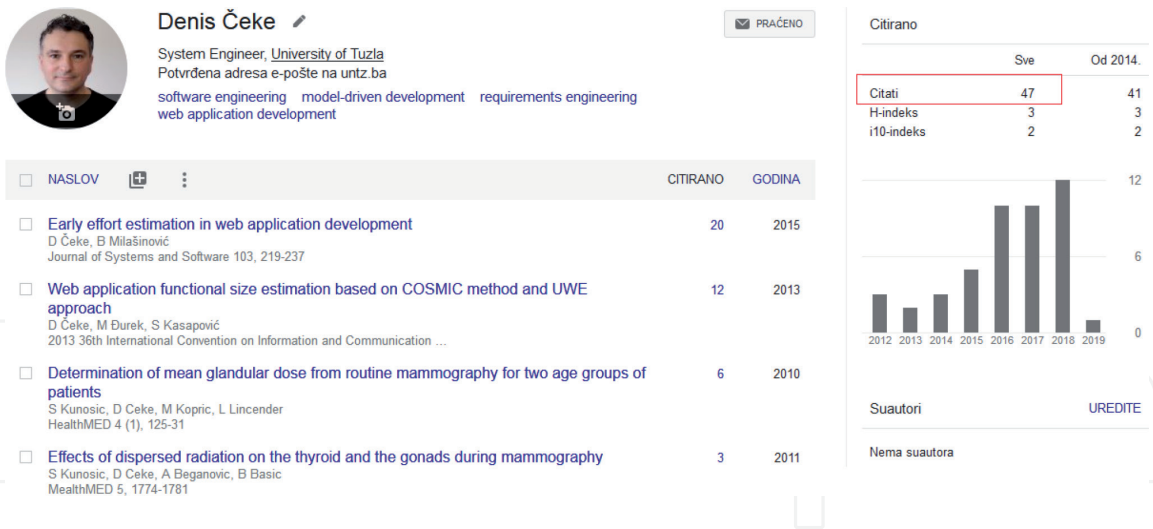


Figure 3.  
Author's profile on GS with an accurate number of quotes.



Figure 4.  
Author's profile on GS after adding two papers with belonging quotes.

which do not belong were ascribed to the mentioned profiles [76]. This data becomes extremely significant if we consider the fact that only 8 scientists from the University of Tuzla have more than 400 quotes (the real number is even less when one takes out nonexistent quotes and scientists who do not work at the University of Tuzla anymore). Analyzing the following 10 profiles of scientists (by order of 11–20), we found 186 more non-belonging quotes [76]. This becomes a greater problem if we bear in mind that most of those “added” works were published in journals of extremely good quality [76], so the quotes have been distributed over the years which affects the parameters of ranking of the university even more. If we know that the first ten ranked profiles are taken into consideration in ranking, we have to ask ourselves if those are really the best ranked scientists and if the position of the university on the Webometrics list is dully calculated [76]. A checkup may easily show that the order, the number of quotes, and the index are not correct, which leads us to a process of incorrect ranking of profiles and incorrect evaluation of the Webometrics list. Knowing the evaluation system on which webometrics functions, it is easy to conclude that adding highly quoted papers from prestigious publications can significantly improve the position of a university. This is also extremely important for ranking of scientists because many universities in the world do not have any access to WOS, and it is very easy for all to use GS ranking lists to measure quality of scientists and a degree of their being quoted. These lists



should be reviewed in order to obtain a realistic picture in ranking of scientists in institutions and by states as an adequate hierarchy of universities. The example from our experiment can easily show that h-index has changed significantly. By adding only two papers, it increased from 3 to even 5, and i-index increased from 2 to 4 (**Figures 3 and 4**). Since the mentioned omissions were found at many other universities as well, we believe that changes in the ranking order of universities and scientists would be very significant. Authors on profiles often are not aware that those are not their works because they update their profiles automatically. Many of them do not pay great attention to it, while a great number of scientists do not have a great knowledge about ranking systems. Another disadvantage of this system is retention of scientists' profiles, although they are engaged by another institution. This means that their profiles are retained after cessation of their engagements. Movements of scientists in the system of scientific institutions and going from one institution to another for the purpose of increasing quality of individual institutions are also a process which should be followed up by GS. To solve this problem, one needs to design a system which would obligate scientific institutions to update data in time and to ask for removal of profiles of those scientists who are not engaged anymore (i.e., the profiles should be adjusted to the new institution). Gaining benefits for quotes on profiles of retired scientists and of those who passed away by institutions is not fair. We found such omission at our university where even 2 out of the first 12 scientists left the University of Tuzla, having together over 1300 citations and being involved in the process of ranking of the university [76]. This omission may not be ascribed only to GS, but one needs to design a system of annual verification of profiles for which scientific institution would be in charge. On the other hand, there are scientists who are not ranked within the frame of an institution due to wrong entries of affiliation and do not contribute to the reputation of a university although they have a great number of quotes. Although Webometrics in its rules hinted at a possibility of sanctioning institutions for double profiles and ascribing papers to wrong authors [60], we noticed that the mentioned sanction has not been implemented, and we saw such cases at several universities in Bosnia and Herzegovina. There is a simple way to solve the mentioned manipulations and omissions. One of the ways is introduction of Z score system [77–79] into Webometrics ranking system. Namely, this system would check authorship, i.e., it would be impossible that a system or a man adds a paper to a scientist's profile if he/she is not its author. For the purpose of further increase of quality of ranking related to publications, one could think about introduction of a certain percentage of ranking on the basis of publications indexed in the best databases in WOS through this system. Authors would also be ranked by the volume of work, i.e., Z score would perform ranking of scientist on the basis of a type of authorship (first author, corresponding author, other authors), number of authors of a paper, quality of a journal, and a number of quotes. Should this process be too demanding in the first phase, a process of filtration of profiles and removal of double profiles, non-belonging papers, and other omissions could be performed in the first phase.

## 5. Conclusion

As a global ranking system, webometrics represents an important step in the assessment of scientific institutions and scientists. It is very important, especially in countries which do not implement international standards and criteria through their institutional educational system. Through its main four parameters, the system entices healthy competition among scientific institutions and scientists. Along with all of the advantages, the system, like any other system, has some



disadvantages, mostly in the domain of ranking of scientists and therefore of institutions through valorization of publications and quotes. Manipulations with a number of quotes and calculation of  $h$ -index and  $i$ -index can be removed with the application of new systems for measuring, such as  $Z$  score, or by introduction of new algorithms for recognition and prevention of these disadvantages.

### Author details

Suad Kunosic<sup>1\*</sup>, Denis Čeke<sup>2</sup> and Enver Zerem<sup>3</sup>

1 Department of Physics, Faculty of Natural Sciences and Mathematics, University of Tuzla, Tuzla, Bosnia and Herzegovina

2 University of Tuzla, Tuzla, Bosnia and Herzegovina

3 Department of Medical Sciences, Academy of Sciences and Arts of Bosnia and Herzegovina, University Džemal Bijedić Mostar, Bosnia and Herzegovina

\*Address all correspondence to: suad.kunosic@untz.ba

### IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Björneborn L, Ingwersen P. Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*. 2004;**55**:1216-1227. DOI: 10.1002/asi.20077
- [2] Thelwall M. A history of webometrics. *Bulletin of the American Society for Information Science and Technology*. 2012;**38**:18-23. DOI: 10.1002/bult.2012.1720380606
- [3] Almind TC, Ingwersen P. Informetric analyses on the world wide web: Methodological approaches to 'webometrics'. *Journal of Documentation*. 1997;**53**:404-426. DOI: 10.1108/EUM0000000007205
- [4] Björneborn L, Ingwersen P. Perspective of webometrics. *Scientometrics*. 2001;**50**:65-82. DOI: 10.1023/A:1005642218907
- [5] Thelwall M. Introduction to Webometrics: Quantitative web research for the social sciences. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 2009;**1**:1-116. DOI: 10.2200/S00176ED1V01Y200903ICR004
- [6] Ingwersen P. The calculation of web impact factors. *Journal of Documentation*. 1998;**54**:236-243. DOI: 10.1108/EUM0000000007167
- [7] Åström F. Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990-2004. *Journal of the American Society for Information Science and Technology*. 2007;**58**:947-957. DOI: 10.1002/asi.20567
- [8] Zhao. Information science during the first decade of the web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology*. Wiley Online Library n.d. 2008. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/asi.20799> [Accessed: January 30, 2019]
- [9] Thelwall M. Extracting macroscopic information from web links. *Journal of the American Society for Information Science and Technology*. 2001;**52**:1157-1168. DOI: 10.1002/asi.1182
- [10] Bar-Ilan J. The use of web search engines in information science research. *Annual Review of Information Science and Technology*. 2004;**38**:231-288. DOI: 10.1002/aris.1440380106
- [11] Aguillo IF, Granadino B, Ortega JL, Prieto JA. Scientific research activity and communication measured with cybermetrics indicators. *Journal of the American Society for Information Science and Technology*. 2006;**57**:1296-1302. DOI: 10.1002/asi.20433
- [12] Thelwall M. Webometrics: Emergent or doomed? n.d. Available from: <http://www.informationr.net/ir/15-4/colis713.html> [Accessed: January 28, 2019]
- [13] Li X, Thelwall M, Musgrove P, Wilkinson D. The relationship between the WIFs or inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*. 2003;**57**:239-255. DOI: 10.1023/A:1024189702463
- [14] Thelwall M, Stuart D. The use of Webometrics for the analysis of knowledge flows within the European research area. Framework Service Contract for Expert Support with the Production and Analysis of R&D Policy Indicators. 2009. Available from: [http://ec.europa.eu/invest-in-research/pdf/download\\_en/spa6\\_final\\_report.pdf](http://ec.europa.eu/invest-in-research/pdf/download_en/spa6_final_report.pdf) [Accessed: January 28, 2019]
- [15] Thelwall M. Bibliometrics to webometrics. *Journal of Information*

Science. 2008;**34**:605-621. DOI: 10.1177/0165551507087238

[16] Thomas O, Willett P. Webometric analysis of departments of librarianship and information science. *Journal of Information Science*. 2000;**26**:421-428. DOI: 10.1177/016555150002600605

[17] Barjak F, Thelwall M. A statistical analysis of the web presences of European life sciences research teams. *Journal of the American Society for Information Science and Technology*. 2008;**59**:628-643. DOI: 10.1002/asi.20776

[18] Barjak F, Li X, Thelwall M. Which factors explain the web impact of scientists' personal homepages? *Journal of the American Society for Information Science and Technology*. 2007;**58**:200-211. DOI: 10.1002/asi.20476

[19] Thelwall M, Vaughan L, Björneborn L. Webometrics. *Annual Review of Information Science and Technology*. 2005;**39**:81-135. DOI: 10.1002/aris.1440390110

[20] Yang B, Qin J. Data collection system for link analysis. In: Pichappan P, Abraham A, editors. *Third IEEE International Conference on Digital Information Management (ICDIM)*, November 13-16, 2008, London, UK, Proceedings. IEEE. 2008. pp. 247-252. DOI: 10.1109/ICDIM.2008.4746781

[21] SocSciBot: Link crawler for the social sciences. n.d. Available from: <http://socscibot.wlv.ac.uk/> [Accessed: January 29, 2019]

[22] Webometric Analyst: Altmetrics, web citation analysis, alternative indicators, link analysis, social web. n.d. Available from: <http://lexiurl.wlv.ac.uk/> [Accessed: January 29, 2019]

[23] Thelwall M, Sud P. Webometric research with the Bing search API 2.0.

*Journal of Informetrics*. 2012;**6**:44-52. DOI: 10.1016/j.joi.2011.10.002

[24] Sherman C, Price G, Sullivan D. *The Invisible Web: Uncovering Information Sources Search Engines Can't See*. Information Today, Inc./CyberAge Books; 2001

[25] Gomes B, Smith BT. Detecting query-specific duplicate documents. US6615209B12003

[26] Thelwall M. Extracting accurate and complete results from search engines: Case study windows live. *Journal of the American Society for Information Science and Technology*. 2008;**59**:38-50. DOI: 10.1002/asi.20704

[27] Vaughan L, Thelwall M. Search engine coverage bias: Evidence and possible causes. *Information Processing and Management*. 2004;**40**:693-707. DOI: 10.1016/S0306-4573(03)00063-3

[28] Thelwall M. Quantitative comparisons of search engine results. *Journal of the American Society for Information Science and Technology*. 2008;**59**:1702-1710. DOI: 10.1002/asi.20834

[29] LexiURL Searcher Classic. n.d. Available from: <http://lexiurl.wlv.ac.uk/searcher/usingLexiURLSearcherClassic.html> [Accessed: January 30, 2019]

[30] Program Package Pajek/PajekXXL n.d. Available from: <http://mrvar.fdv.uni-lj.si/pajek/> [Accessed: January 29, 2019]

[31] UCINET Software n.d. Available from: <https://sites.google.com/site/ucinetsoftware/home> [Accessed: January 29, 2019]

[32] Borgatti S. NetDraw 2002. Available from: <https://sites.google.com/site/netdrawsoftware/home> [Accessed: January 29, 2019]

- [33] Cooper K. Keyword research, competitor analysis, & website ranking | Alexa. AlexaCom n.d. Available from: <https://www.alexa.com/> [Accessed: January 29, 2019]
- [34] Alexa Internet. Firefox Browser Extension Download. n.d. Available from: <https://www.alexa.com/toolbar> [Accessed: January 29, 2019]
- [35] WHOIS Search, Domain Name, Website, and IP Tools—Who.is n.d. Available from: <https://who.is/> [Accessed: January 29, 2019]
- [36] SEO Tools—Search Engine Optimization Tools. n.d. Available from: <https://www.webconfs.com/> [Accessed: January 29, 2019]
- [37] Majestic®: Marketing Search Engine and SEO Backlink Checker n.d. Available from: <https://majestic.com/> [Accessed: January 29, 2019]
- [38] The ultimate SEO Research Tool | Searchmetrics Essentials. Searchmetrics 2016. <https://www.searchmetrics.com/essentials/> [Accessed: January 29, 2019]
- [39] Google PageRank Checker—Check Google page rank instantly n.d. Available from: [https://www.prchecker.info/check\\_page\\_rank.php](https://www.prchecker.info/check_page_rank.php) [Accessed: January 29, 2019]
- [40] Ahrefs Team. n.d. Available from: <https://ahrefs.com/about> [Accessed: January 29, 2019]
- [41] Site Explorer by Ahrefs: Competitive Analysis Tool For SEO. n.d. Available from: <https://ahrefs.com/site-explorer> [Accessed: January 29, 2019]
- [42] Content Explorer by Ahrefs: Get Content Ideas on Any Topic. n.d. Available from: <https://ahrefs.com/content-explorer> [Accessed: January 29, 2019]
- [43] Keywords Explorer by Ahrefs: A Potent Keyword Tool. n.d. Available from: <https://ahrefs.com/keywords-explorer> [Accessed: January 30, 2019]
- [44] Rank Tracker by Ahrefs: Check & Track Keyword Rankings. n.d. Available from: <https://ahrefs.com/rank-tracker> [Accessed: January 29, 2019]
- [45] SEO Site Audit tool by Ahrefs. n.d. Available from: <https://ahrefs.com/site-audit> [Accessed: January 29, 2019]
- [46] Definition of RANKING n.d. Available from: <https://www.merriam-webster.com/dictionary/ranking> [Accessed: January 29, 2019]
- [47] Vernon MM, Balas EA, Momani S. Are university rankings useful to improve research? A systematic review. *PLoS One*. 2018;**13**:e0193762. DOI: 10.1371/journal.pone.0193762
- [48] Aguillo IF, Bar-Ilan J, Levene M, Ortega JL. Comparing university rankings. *Scientometrics*. 2010;**85**: 243-256. DOI: 10.1007/s11192-010-0190-z
- [49] Google. n.d. Available from: <https://www.google.com/> [Accessed: January 30, 2019]
- [50] Consejo Superior de Investigaciones Científicas (CSIC). csic.es. n.d. Available from: <http://www.csic.es/> [Accessed: January 29, 2019]
- [51] Aguillo IF, Ortega JL, Fernández M. Webometric ranking of world universities: Introduction, methodology, and future developments. *Higher Education in Europe*. 2008;**33**:233-244. DOI: 10.1080/03797720802254031
- [52] ARWU World University Rankings 2018 | Academic Ranking of World Universities n.d. Available from: <http://www.shanghairanking.com/> [Accessed: January 29, 2019]
- [53] Wouters P, Reddy C, Aguillo I. On the visibility of information on the web:



An exploratory experimental approach. *Research Evaluation*. 2006;**15**:107-115. DOI: 10.3152/147154406781775931

[54] Web indicators: A new generation of S&T indicators › KNAW Research Portal. n.d. Available from: [https://pure.knaw.nl/portal/en/publications/webindicators-a-new-generation-of-st-indicators\(3a0f58a7-14dc-47c5-9123-ee3c8b0983d5\)/export.html](https://pure.knaw.nl/portal/en/publications/webindicators-a-new-generation-of-st-indicators(3a0f58a7-14dc-47c5-9123-ee3c8b0983d5)/export.html) [Accessed: January 29, 2019]

[55] Bar-Ilan J. Expectations Versus Reality—Search Engine Features Needed for Web Research at Mid 2005; 2005

[56] Link Analysis: An Information Science Approach. Vol. 04. Emerald Group Publishing Limited; 2004. pp. 265-269. DOI: 10.1108/S1876-0562(2004)0000004028

[57] Cothey V. Web-crawling reliability. *Journal of the American Society for Information Science and Technology*. 2004;**55**:1228-1238. DOI: 10.1002/asi.20078

[58] Ranking Web of Universities: More than 28000 institutions ranked. Updated every January and July: Webometrics, Bibliometrics, Altmetrics. n.d. Available from: <http://www.webometrics.info/en> [Accessed: January 29, 2019]

[59] Zitt M, Filliatreau G. Big Is (Made) Beautiful some Comments about the Shanghai Ranking of World-Class Universities. 2007. pp. 141-160

[60] Methodology. Ranking Web of Universities: More than 28000 institutions ranked. n.d. Available from: <http://www.webometrics.info/en/Methodology> [Accessed: January 29, 2019]

[61] Zerem E. Right criteria for academia in Bosnia and Herzegovina. *The Lancet*. 2013;**382**(9887):128

[62] Zerem E. Academic community in a developing country: Bosnian realities. *Medicinski Glasnik (Zenica)*. 2014;**11**(2):248-251

[63] Ranking Web of Universities: Ranking by Areas. Available from: <http://www.webometrics.info/en/node/18> [Accessed: April 12, 2019]

[64] Ranking Web of Universities: Bosnia and Herzegovina. Available from: <http://www.webometrics.info/en/Europe/Bosnia%20and%20Herzegovina%20> [Accessed: April 12, 2019]

[65] Constitutional Court of Bosnia and Herzegovina: Decisions. Available from: [http://www.ustavnisud.ba/dokumenti/\\_bs/U-19-16-1067890.pdf](http://www.ustavnisud.ba/dokumenti/_bs/U-19-16-1067890.pdf) [Accessed: April 12, 2019]

[66] Constitutional Court of Bosnia and Herzegovina: The Case in the Prosc. Available from: <http://www.ccbh.ba/odluke/#>

[67] Butler D. Computing giants launch free science metrics. *Nature*. 2011;**476**(7358):18-18

[68] Jacso P. Deflated, inflated and phantom citation counts. *Online Information Review*. 2006;**30**(3):297-309. [Accessed: April 12, 2019]

[69] Jacsó P. The plausibility of computing the h-index of scholarly productivity and impact using reference-enhanced databases. *Online Information Review*. 2008;**32**(2):266-283

[70] Jacsó P. Metadata mega mess in Google Scholar. *Online Information Review*. 2010;**34**(1):175-191

[71] Jacsó P. Calculating the h-index and other bibliometric and scientometric indicators from Google Scholar with the publish or perish software. *Online Information Review*. 2009;**33**(6):1189-1200

[72] Aguillo IF. Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*. 2012;**91**(2):343-351

[73] Fong EA, Wilhite AW. Authorship and citation manipulation in academic research. *PLoS One*. 2017;**12**(12):e0187394

[74] Lopez-Cozar ED, Robinson-Garcia N, Torres-Salinas D. Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting. In: *EC3 Working Papers 6*. 2012

[75] Delgado-López-Cózar E, Cabezas-Clavijo Á. Google Scholar Metrics: An unreliable tool for assessing scientific journals. *El Profesional de la información*. 2012;**21**(4):1386-6710

[76] Google Scholar: Profiles University of Tuzla. Available from: [https://scholar.google.hr/citations?view\\_op=view\\_org&hl=hr&org=13570989470062067060](https://scholar.google.hr/citations?view_op=view_org&hl=hr&org=13570989470062067060) [Accessed: April 7, 2019]

[77] Zerem E. The ranking of scientists based on scientific publications assessment. *Journal of Biomedical Informatics*. 2017;**75**:107-109

[78] Zerem E. The ranking of scientists. *Journal of Biomedical Informatics*. 2018;**79**:147

[79] Zerem E, Kunosić S. The ranking of scientists: Computational calculation of Z-score. *Journal of Biomedical Informatics*. 2018;**81**:133