International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 7 No. 4 Aug. 2014, pp. 1330-1337

http://www.ijias.issr-journals.org/

Global University Ranking System: A New Approach by Combining Academic Performance and Web-Based Indicators using Clustering

Abbas Fadhil Mohammed Ali AL-Juboori

Department of Computer Science, College of Science, University of Kerbala,
Kerbala, Iraq

ABSTRACT: Global university rankings have cemented the notion of a world university market arranged in a single “league table” for comparative purposes and have given a powerful impetus to intranational and international competitive pressures in the sector.
The studies on evaluation of academic productivity and quality and web presence have led to development of new academic fields such as Bibliometrics, Scientometrics, Informetrics and Webometrics.
During the last thirty years, as an outcome of these new emerging academic fields, several university ranking systems have been developed both at national and global level. Although these university ranking systems have attracted attention; they have been criticized due to a number of issues such as inappropriateness of indicators chosen, scoring procedure adopted, weighting, etc. In this study, a combined academic and web performance evaluation and global ranking system has been developed and implemented using data mining. The new global system which processes about more than 5,000 world universities is based on data from non-subjective, reliable and universally accepted online sources. The scoring procedure includes statistical analysis and data has been collected via a tool developed for this purpose to eliminate human errors.

KEYWORDS: Ranking, Web performance, Academic performance, League tables, Web.

1 INTRODUCTION

University ranking tables are a global phenomenon [1] with more than 30 years of history. Rankings began in 1983, when the US News and World report started to publish the annual America’s best colleges review. It spawned the development and publication of more and more ranking tables from numerous countries across the world. Over the last two decades, higher education ranking tables have emerged not only from the private and media-based sectors, but also from professional associations and governments [2]. The goals of ranking tables include: (1) directing an entrant to higher educational programmes, (2) evaluating the phenomena of the international higher education market, (3) introducing market directions for universities at national levels, (4) enhancing sound and positive competition for students, professors, and the funders of universities [3]. Ranking tables also offer information about the quality and other characteristics of higher education institutions, influencing the students’ matriculation. In many countries higher education presents a financial burden for students, their parents, and scholarship foundations. When students are granted a scholarship, it is important for them to receive high quality education and other complementary services, thus influencing their future employment possibilities.

There is much criticism about ranking methodologies. Marginson [4] argues that ranking tables conceal a whole array of methodological problems (e.g. weighing) and anomalies, regarding the indicators. University ranking requires a specific definition and quality criteria for indicators developed in order to assess a university’s performance. As a result of methodological problems and usage of various indicators, overall rankings and ranking tables differ. It is often unclear why a particular methodology or indicator was chosen, how well it was founded, by whom it was decided and how open and reflective the decision process was [5]. Regarding the indicators, the Academic Ranking of World Universities (ARWU) ranking table uses objective data, measured quantitatively, but the winners are a particular kind of science-oriented universities.

Times Higher Education Supplement (THES) ranking table relies heavily on subjective evaluations by experts and recruiters, directed towards the prestige and power of the university. Marginson [4] argues that the THES ranking table does not express the quality of education, and only slightly touches the research dimension. The latter merely takes account of the number of citations per faculty member, which contributes 20% to the total ranking score.

Regarding the methodologies, data acquisition presents another problem of university rankings. Many ranking tables are based on subjective, qualitative data, received from universities themselves. In the THES ranking table, 50% of the final score is based on subjective evaluations by peer reviewers. Researchers at the Leiden University carried out a correlation between rankings based on subjective evaluations, and citations counting an accepted measure of scientific impact. The results showed that there was no correlation ($R^2 = 0.005$) between them [6]. This has fostered doubts about the credibility of the THES ranking table [7]. Doubts exist regarding the US News ranking table, too. The Institute of Educational Policy from Toronto documented evidence of cheating by universities listed in the US News [7]. Disagreement with the ranking methodologies makes universities hesitate to participate in the ranking.

Nowadays, universities are exhaustively compared from the educational and research perspective, such as student to staff ratio, number of citations, or number of scientific publications. In contrast, the implication of environmental issues has received little or no attention, although many universities are monitoring their environmental footprints.

In the light of the weaknesses of existing ranking systems, a new ranking system has been designed to rank world universities by using indicators which measures academic productivity and quality, teaching quality, knowledge transfer, and Web performance. The study is based on non-subjective and confirmed sources and it is supported by statistical studies in order to apply fair scoring procedure. Moreover, in this study, an automated data collection tool has been developed and implemented to lessen human oriented errors in data collection and a great effort has been spent for institution naming.

The history of higher education rankings goes back to 1983. In that year, Bob Morse, from US News and World Report, published "American Colleges" ranking. However, it began to be published annually in the year 1987. By following US News and World Report ranking, many national and global ranking systems have been developed [8].

2 A NEW SYSTEM FOR GLOBAL RANKING METHODOLOGY

GURS is a ranking system designed to process data of about 5,000 HEIs from all around the world in order to evaluate them according to various indicators. It bases on confirmed, objective and reliable sources. Many processes from data collection to scoring are conducted programmatically. Statistical studies are another main part of the project since they have a crucial role on scoring procedure and result analysis.

The basic steps of this ranking system are given in fig. 1.

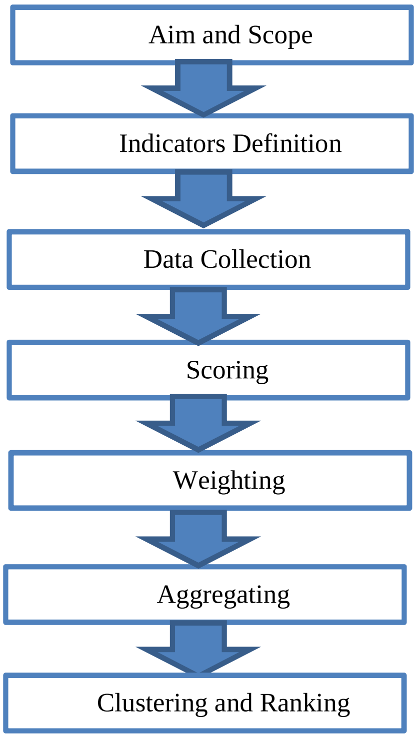

Fig. 1. The Basic Steps

2.1 GURS Indicators

A summary of indicators is displayed in table 1 below;

Table 1. GURS indicators

No.	Indicators	Objectives	Years	Sources	Weight
1.	No. of Papers	Research Productivity	2009-2013	Scopus and Google Scholar	15%
2.	No. of Citations	Research Quality	2009-2013	Scopus and Google Scholar	20%
3.	No. of Patents	Knowledge Transfer	2009-2013	Scopus and Google Scholar	5%
4.	Student/Staff Ratio	Teaching Quality	2012	Institutions website, Government ministries, Agencies and Wikipedia	10%
5.	Research Centers	Research Quantity	2013	Institutions website, Government ministries, Agencies and Wikipedia	10%
Web-based Indicators					
6.	No. of Backlinks	Visibility	2013	Majestic SEO and Ahrefs	10%
7.	No. of Referring Domains	Impact	2013	Majestic SEO and Ahrefs	10%
8.	Alexa Traffic Rank	Website Popularity	2013	Alexa	10%
9.	Website Reputation	Reputation	2013	Alexa	10%

2.2 Data Collection Module

The data collection system heavily bases on automatic processes in order to eliminate human based errors and accomplish the study in a shorter time period. Thus, Microsoft Office Excel 2010, Quick Macro (QMacro) 6.6 Scripting Edition and Visual Basic for Applications (VBA) tools are used to develop an automatic collection program and collect data.

3 STATISTICAL ANALYSIS AND CLUSTERING

All of the analysis will be based on the raw data gathered for the indicators. In order to accomplish statistical analysis Microsoft Office Excel 2010 and IBM SPSS software was used.

3.1 DESCRIPTIVE STATISTICS

The descriptive statistics provides quantitative descriptions of indicators as presented in the table 2 Descriptive Statistics. The sample of this analysis consists of 5000 world universities selected according to criteria defined previously. The selected sample‘s raw data are analyzed in nine categories, namely indicators.

Table 2. Descriptive Statistics of the indicators

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness
Papers	5000	1	288000	14229.751	28021.379	785197695.4	5.028
Citations	5000	0	112000	1578.956	6040.303	36485260.54	8.672
Patents	5000	0	63000	507.685	2407.712	5798241.143	11.561
S/S Ratio	5000	2	150	16.959	11.410	130.236	4.747
Res. Cen.	5000	0	112	26.401	16.642	276.972	0.692
Backlinks	5000	1011	7349603	281607.943	403017.595	1.62423E+11	7.260
Domains	5000	11	661670	3466.631	15106.770	228214515.4	30.246
Alexa T. R.	5000	23	37491084	56248.935	555229.306	3.0828E+11	61.895
Reputation	5000	2	66890	3008.615	4170.941	17396750.49	5.083

The indicators‘ central tendencies are expressed in terms of means. According to analysis the average paper number of 5000 universities is 14229.751 and the average number of citations received for the papers published in years between 2009 and 2013 is 1578.956. Patent average number for the five years was 507.685. The teaching quality average number for all universities which is represented by student/staff ratio has value of 16.959 and for research center was 26.401. For the web-based indicators backlinks, domains, Alexa traffic rank and web site reputation were 281607.943, 3466.631, 56248.935 and 3008.615 respectively.

The spread of data for each indicator can be evaluated via standard deviation values. In fact, the minimum and maximum values of indicators provide a view on the wide dispersion of values. As an example, the raw data of papers indicator vary between 1 and 288000 Standard deviation of the same indicator is 28021.379. It means that the data is spread out over 28021.379 from the mean which is 14229.751.

This large range also occurs in other indicators. It is due to the size variance of selected universities. In other words, since there are 5000 universities from all around the world, their published documents, received citations, patents, s/s ratio and research centers are expected to be varied in a wide range. It is same with the web-based indicators.

4 CLUSTERING

In order to decrease the best university‘s effect on remaining ones ‘scores, and to classify the institutions into 5 levels, data mining is applied by using K-Means algorithm to cluster the total score into 5 groups. IBM SPSS Clementine software used to implement this algorithm. The clusters, their records, mean and the standard deviation for each cluster are given in table 3 below:

Table 3. Clustering of GURS

Cluster	Number of Institutions	Mean	Std. Deviation
1	14	93.624	3.743
2	3464	41.204	1.502
3	1157	68.781	1.434
4	446	48.961	3.522
5	30	79.808	3.954

It is clear that the cluster 1 is the highest level group of 14 institutions which has 93.624 average number of total score , while the lowest level represented by cluster 2 containing 3464 institutions with average number of total score 41.204.

The two screen shots of Clementine software in fig. 2 and fig. 3 below explained the Pei charts and clusters ratio using K-Means algorithm and the frequency distribution of the total scores.

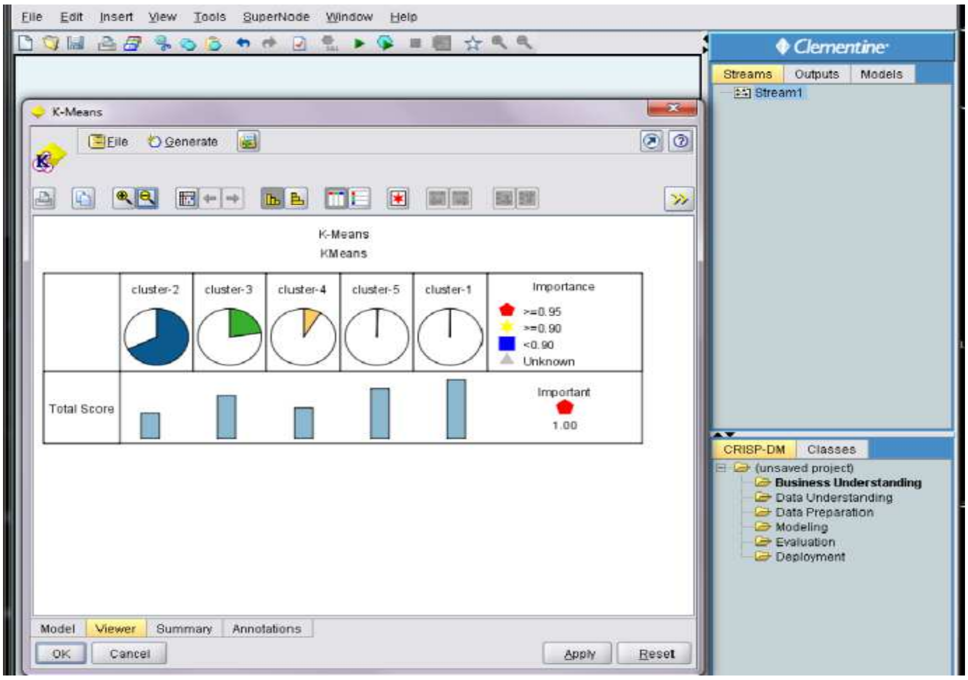

Fig. 2. The Screen shot of clustering result using K-Means algorithm

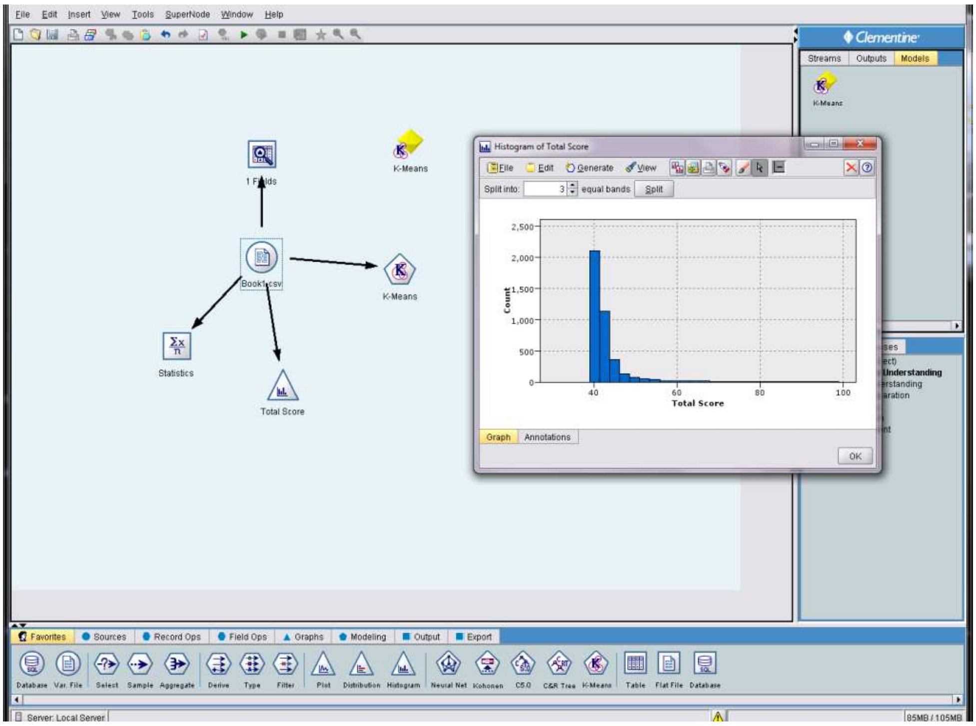

Fig. 3. The Screen shot of Total Score Distribution

5 RANKING RESULTS

After completion of data collection, statistical analysis, scoring procedures and clustering, a new list containing indicator based scores and overall score of each university has been created. The sorting of overall scores in descending order in this list produces the ranking of Top 5,000 world universities. According to the results, the Top 50 of them is as in table 4.

Table 4. Top 50 universities and Total Scores

Rank	University	Total Score	Rank	University	Total Score
1	Harvard University	98.545	26	Cornell University	80.155
2	Stanford University	97.914	27	University of California Berkeley	80.154
3	University of Pennsylvania	96.374	28	State University of New York	80.142
4	University of North Carolina	96.045	29	Barry University	80.044
5	University of Michigan	95.883	30	University of Hong Kong	79.791
6	New York University	94.306	31	University of North Texas	79.676
7	University of Oxford	94.039	32	University of Virginia	78.503
8	University College London	92.785	33	Wright State University	78.203
9	Massachusetts Institute of Technology	92.565	34	University of New Mexico	77.706
10	George Washington University	92.214	35	Long Island University	76.004
11	Columbia University New York	90.409	36	University of Tokyo	75.715
12	Michigan State University	90.269	37	University of Bath	75.640
13	University of Toronto	90.141	38	Johns Hopkins University	75.063
14	University of Florida	88.687	39	University of Minnesota	75.033
15	Princeton University	88.106	40	Indiana University	75.024
16	Carnegie Mellon University	87.068	41	Rice University	74.770
17	University of Cambridge	86.626	42	Boston University	74.486
18	Duke University	86.270	43	National Taiwan University	74.383
19	University of Southern California	85.696	44	Florida State University	73.875
20	Yale University	85.327	45	University of Edinburgh	73.739
21	University of Western Australia	83.660	46	York University	73.206
22	University of Georgia	82.074	47	Northwest University	72.959
23	University of Chicago	81.046	48	Australian National University	72.840
24	University of Washington	80.369	49	American University Washington DC	72.010
25	Ohio University	80.350	50	Central Washington University	71.881

6 DISCUSSIONS ON RANKING RESULTS

6.1 TOTAL SCORES DISTRIBUTION

Total scores distribution versus ranking of universities are given in fig. 4 Total scores distribution. The total scores of institutions are distributed in a non linear trend line against their corresponding ranks. There is a rapid increase in total score for the first left most tail and rapid decrease for the first most right tail. The reason of the increasing is the effect of the outlier which is the best performing universities in the list, while the reason of the decreasing is worst performing universities in the list.

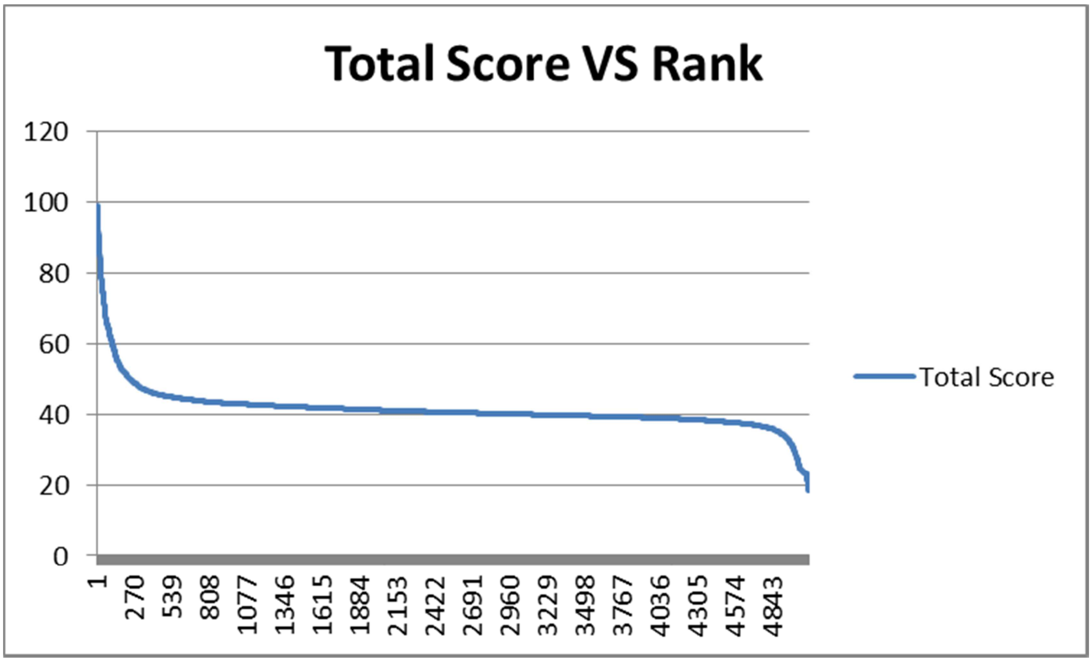

Fig. 4. Total scores distribution

6.2 GURS VERSUS OTHER RANKING SYSTEMS

The Coefficient of Determination (R^2) has been evaluated for each comparison to indicate the similarity ratio between any two ranking systems.

In the table 5 below, R^2 values are given for each comparison. The results indicate that the highest similarity has been found in comparison of WEBOMETRICS due to the nature of its web-based indicators and the big number of ranked universities like our system. The lowest one is ARWU due to the nature of its ranking methodology. Another significant result is that GURS is more similar to HEEACT ranking than QS ranking.

Table 5. Comparison with ranking systems

	ARWU	URAP	WEBOMETRICS	QS	HEEACT
GURS	60.90%	69.40%	81.50%	77.30%	79.30%

Another comparison of ranking systems will be on the universities. The table 6 provides some university examples which are in the Top 500 of GURS and their corresponding rank numbers in the systems.

Table 6. University examples in GURS and other ranking systems

Universities	GURS	ARWU	URAP	Webometrics	QS	HEEACT
New York University	6	27	52	35	43	37
Duke University	18	36	20	17	20	18
University of Tokyo	36	20	10	48	30	16
Universität Wien	237	151-200	169	84	160	202
Seoul National University	332	101-150	39	281	37	69
Cardiff University	346	151-200	152	484	143	147
Université de Geneve	435	69	160	137	74	117

7 Conclusions And Future Works

In this study, a system, names as GURS, designed for ranking Top 5,000 world universities according to their both academic and web performance. GURS is based on non subjective and confirmed sources. In the scope of this study, a new tool has been developed to collect data from various sources. Moreover, statistical analyses on scoring procedure have been conducted. Data mining technique using K-Means algorithm was applied to cluster the universities into 5 groups levels in terms of their total scores.

To sum up, this study can support improvement of institutions academically and their web sites as well by evaluating their current situation. It is especially important for the universities of emerging and developing countries. The increase in academic quality and web visibility of HEIs supports the scientific development. It contributes to prosperity of the country. Thus, worldwide prosperity and peace can be settled.

There are issues which have been left out of the scope in this study for future studies. Firstly, the methodology of the current ranking system may be developed by adding new indicators. Moreover, ranking institutions by their subject areas, fields or disciplines may be another future study. Thus, the institutions might be compared according to their specialized areas instead of as a whole. The number of institutions which have been processed and ranked is about 5000. In the future studies, thousands of new institutions may be added into the list.

There may be future studies on data collection process by developing more user friendly tools. Another future study on data collection can be obtaining required data from databases directly instead of limited web interface.

References

[1] Berlin Principles Berlin Principles. Available online: http://www.cepes.ro/hed/meetings/berlin06/Berlin%20Principles.pdf; 2006 [6 November 2006].

[2] UNESCO CEPES, European Centre for Higher Education. *Higher education ranking systems and methodologies: how they work, what they do*. Available online: http://www.cepes.ro/hed/policy/ranking.htm [13 January 2006].

[3] Jesensek M. *Why the rankings are important?* Katedra (Maribor) 2006;1(4):6–7 [in Slovene].

[4] Marginson S. *Global university rankings: implications in general and for Australia*. Journal of Higher Education Policy and Management 2007;29(2):131–42.

[5] RFSU – The Ranking Forum of Swiss *Universities. Ranking methodology: how are universities ranked?* Available online: http://www.universityrankings.ch/en; 2008

[6] Van Raan AFJ. 2005. *Challenges in ranking of Universities*. Invited paper for the First International Conference on World Class Universities, Shanghai Jaio Tong University, Shanghai, June 16–18, 2005.

[7] Enserink M. *Who ranks the university rankers?* Science 2007;317:1026–8.

[8] Jarocka M. (2012). *University Ranking Systems-From League tables to Homogeneous Groups of Universities.* World Academy of Science, Engineering and Technology 66 2012.

[9] http://ahrefs.com