The Social Sciences 13 (3): 763-765, 2018
ISSN: 1818-5800

University Webometrics Ranking using Multicriteria Decision Analysis: Entropy and TOPSIS Method

Handaru Jati
Department of Informatics Engineering Education,
Universitas Negeri Yogyakarta, Karangmalang, Indonesia

Abstract: For many academic institutions, among which are the universities the web has become an interesting tool. This web presence of academic institutions has led the researcher, academic and scientific publication in this environment to use this web to reflect their activities. This study explores the webometrics ranking for world universities. The webometrics for world universities were calculated by using two quantitative techniques in Multicriteria Decision Analysis (MCDA) which are Entropy method and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). This calculation was made based on four key indices: size, the visibility of the website, rich content size which is the volume of published information and scholar. The basic principle of the TOPSIS method is that the chosen alternative should have the "shortest distance" from the ideal solution and the "farthest distance" from the "negative-ideal" solution. For the case studies investigated the entropy and TOPSIS technique is effective and simple in terms of computational implementation. Its advantages is that the algorithm does not require tuning any parameters. These models efficiently help evaluators to determine with a strategic view for future developments and more aspect by using multicriteria decision analysis. It concludes by acknowledging that webometrics ranking systems are viewed differently by different stakeholders and hence can be approached in different ways.

Key words: Ranking, multicriteria, webometrics, MCDA, entropy, TOPSIS

INTRODUCTION

After the advent of Information and Communication Technologies (ICTs), particularly the web is experienced-one was called "information society" because of the range of information available daily to society. The web has been one of today's biggest information sources. It is characterized as one of the greatest intellectuals deposits of society and before this assertion we can see ever more intensely concern about the use, organization and evaluation of this area which has been a priority issue for scholars of information systems. In this context, a new area of study called webometrics. This aims to quantitatively analyze the information available on the web through view variables of sites. This webometrics makes use of tools such as algorithms, search engines, directories, among others as a means of obtaining quantitative data which when analyzed by webometrics indicators define the presence of communication and scientific information on the web. From these indicators we can establish a ranking of the sites which can also be called URL and based on this analysis can indicate which sites stand out for the presence of the same network, focusing on its national and international recognition before other institutions. Such recognition can be characterized from the integrity, availability and reliability of the information provided by these URL's through your links. Seen the growing presence of webometrics in scientific context is that aroused the interest of investigating this method and its applications in order to identify what universities in the world stand out in the web space with primacy to provide information to the virtual scientific academia and also check how they are behaving scientific information in virtual academia and figure out how to webometrics can collaborate in the process of measuring the amount of scientific information and communications available on the network, awakening, hence the interest of improving the quality of this information available. The results to be obtained in research aimed at identifying which university sites, among the selected, stand out in the web space. By achieving these results will be possible to identify which university sites have contributed to the dissemination of scientific information in the virtual space and suggest improvements to the use of this type of metric study the so-called information society and consequently with the information science.

The webometrics quantitatively analyzes the information available in web through study of the links of sites. The webometrics is a scientific discipline that studies the quantitative aspects and information resources on web documents and their use, being based on bibliometric and informetric methods, encompassing 2 basic categories: web link structure and evaluation of search engines using informetric methods (Aguillo *et al*., 2008). The webometrics has stood out increasingly in the scientific field due to the precision of the statistical results of analysis of the information flow to certain web sites. Thus, the sites analyzed collaborate with the completion of the surveys that provide more progress to science and as study and dissemination web support. This method has been widespread in scientific circles, especially after the creation of the webometrics ranking of world universities, an initiative of Cibermetrics Lab., a group of researchers belonging to the Consejo Superior de Investigaciones Cientificas (CSIC) from Spain. These researchers analyze the presence of academic web sites, i.e., all universities world wide web and periodically disclose a ranking of the most recognized universities and connected to the web by other institutions. The webometrics evaluation is to examine between web pages which stand in content structure; use through research and search behavior; quality of the information provided and further through the structure of the existing hypertext sites. From this perspective there is the extreme importance of taking into consideration the publication on the web not only as the main tool for scholarly communication but also as a true reflection of the global organization and consequently the performance of universities (Aguillo *et al*., 2008). From this perspective, there is the importance of analyzing academic sites that has been highlighted in the current scientific scenario, regarding the provision of research and scientific information.

MATERIALS AND METHODS

This research method is based on a quantitative aspect of the world Higher Education Institution sites that have academic and Postgraduate Program listed in http://www.webometrics.info/. Obtaining the list of sites to be analyzed is achieved by using webometrics ranking in 2012. This study uses an approach to data collection: using Google and Bing search engine, offering special search function that matches only elements of web-like pages, domains, inlinks and rich content wise. Collection took place within the same month (February 2012) in order to limit errors associated with frequent updates website. For classification purposes, only those universities or research centers are considered to have independent web domain(s). The visibility is based on link analysis using the external number of incoming links. Three indicators were also added to the website component before qualifying. These are; document number measuring the amount of rich files in a web domain, number of publications that are collected by database Google Scholar, number of websites of each university by Google. The last 20 rows higher universities webometric January 2012 edition were selected for this study. The list of the universities studied is provided in Table 1 and 2. Data sources include: Catalogue of world universities (http://www.webometrics.info/universit-by-country-select.asp.htm). Four indicators obtained from the qualitative results provided by the major search engines are.

Size (S): Number of pages recovered from four engines: Google.

Visibility (V): The total number of unique external links received (inlinks) by a site can be only confidently obtained from Bing.

Rich files (R): Adobe Acrobat (pdf) Adobe PostScript (ps) Microsoft Word (doc) and Microsoft PowerPoint: After evaluation of their relevance to academic and publication activities and considering the volume of the different file formats, the following were selected (.pdf) (.ps), (.ppt).

Scholar (SC): Google Scholar provides the number of study and citations for each academic domain. These results represent data base Academic documents, reports and other academic items. For each engine, results are log-normalized to 1 for the highest value and then combined to generate the rank. Rank/position of a university is ranked is obtained with the help of the following equation:

Webometrics ranking (position) = 4*Rank V+
2**Rank R rows +1+1*Rank Sc

Entrophy method: Entropy method was developed by Zeleny (1982) as an objective method for allocating weights based on the decision matrix without affecting the preference of the decision maker. The relative importance of criterion j in a decision situation, wj measure its weight is directly related to the amount of information provided by the intrinsically set of alternatives with respect to that criterion (Romero and Pomerol, 1997). There is greater diversity in evaluations of the alternatives then greater importance should be the criterion. This diversity is based conceptually on the solid, accepted concept of entropy in an information channel presented by Shannon (1949), the procedure as follows: the evaluations ij (i = 1, m) (j = 1, n) are taken as normalized as

Table 1: List of world universities with the corresponding number of size, visibility, rich files and scholar

			Rich files					
University	Size (k)	Visibility	*.pdf	*.ps	*.ppt	*.doc	Total	Scholar (k)
A	9950	177,321	259000	84200	9110	22900	375210	9950
B	8970	307,113	390000	26400	10800	13400	440600	8970
C	33200	4,616,437	317000	22300	18100	19900	377300	33200
D	30100	362,854	268000	10100	8650	22800	309550	30100
E	26700	113,286	269000	12900	20000	20500	322400	26700
F	31800	144,949	242000	13000	10300	12500	277800	31800
H	5550	415,198	253000	5550	8900	25300	292750	5550
I	13900	138,804	447000	9790	11700	23300	491790	13900
J	776	266,026	135000	2980	8210	10700	156890	776
K	1530	118,231	245000	16700	11100	55500	328300	1530
L	19900	339,721	285000	15500	21400	30300	352200	19900
M	8910	57,035	346000	6090	11600	28100	391790	8910
N	8160	121,089	934000	32000	11000	28600	1005600	8160
O	1930	243,812	202000	10300	9240	20300	241840	1930
P	17000	206,485	140000	7200	6030	9400	162630	17000
Q	5000	73,305	184000	7040	8300	13200	212540	5000
R	6570	151,812	220000	4630	6200	14000	244830	6570
S	8690	709,434	152000	18200	8080	10200	188480	8690
T	14900	660,038	197000	8610	12100	16100	233810	14900
U	3980	90,970	286000	17500	6630	21400	331530	3980

Table 2: List of 5 world universities for determining weight for webometrics criteria

University	Size	Visibility	Rich files	Scholar
A	9950	177,321	375210	9950
B	8970	307,113	440600	8970
C	33200	4.616,437	377300	33200
D	30100	362,854	309550	30100
E	26700	113,286	322400	26700

a fraction of the sum i_{ij} O to the original assessments of each criterion j for m>1 and i = 1, 2, …, m and j = 1, 2, …, n:

$$a_{ij} = \frac{k_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} k_{ij}} \quad (1)$$

Entropy (E_j) is calculated:

$$E_j = \left[\frac{-1}{\ln(m)}\right]\sum_{i=1}^{m}\left[a_{ij}\ln(a_{ij})\right] \quad (2)$$

Where:

m = Number of alternatives in the matrix standardized assessments

i_j = Criteria or standardized attributes

D_j = Diversity criterion is calculated

$$D_j = 1 - E_j \quad (3)$$

The normalized weight of each criterion (W_j) is calculated:

$$W_j = \frac{D_j}{\sum D_j} \quad (4)$$

TOPSIS ranking method: The TOPSIS technique is widely used for solving decision making problems. The TOPSIS method is a decision model proposed order preferences similar to an ideal solution is therefore a method of ranking. It was developed by Hwang and Yoon (1995) also researched by Zeleny (1982), Lai and Hwang (1992) and many more. TOPSIS is a method of multicriteria decision management solutions to identify the s of a finite set of alternatives. The basic principle is that the chosen alternative should have the shortest distance to the positive ideal solution and the greatest distance to the ideal negative solution. An ideal solution is defined as a collection of scores or going Lords in all attributes considered in the decision, it may be that such a solution is unattainable. The compound for the best values of jth attribute on all possible alternatives vector is the one called "positive ideal solution", conversely the "negative ideal solution" is one whose vector contains the worst values in all attributes. The intuitive concept is ideal alternative would be that which without doubt, always choose the decider. Similarly, the anti-ideal alternative would be one that without hesitation, never choose the decider.

Step 1:

- Decide the criteria B_j (j = 1, 2, ….., m where m is the number of criteria/attributes) for selecting the alternative(university websites). The criteria or attributes will be size, visibility, rich file and scholar
- Choose a set of university website alternatives A_i (i = 1, 2, …., n where n is the number of alternative websites considered in the study)

- Measures the performance of each alternative with respect to attributes denoted as X_{ij} (for i = 1, 2, ..., n; j = 1, 2, ..., m)
- Decide the weight or relative importance of each attributes, W_j (j = 1, 2, ..., m)

The values associated with the attributes (x_{ij}) may be in different units. So, the elements of the decision Table 3 are normalized for different alternatives using the following equation:

$$X^*_{ij} X_{ij} / \sum_{i=1}^{n} X_{ij} \quad (5)$$

Where:

X_{ij}^* = The normalized value of X_{ij}

$\sum_{i=1}^{n} x_{ij}$ = The total of the values of jth attribute for 'n' number of alternatives

Step 2: Determine weights of importance of the attributes using result from Entropy method.

Step 3: Find the weighted normalized decision matrix. The weighted normalized value v_{ij} is calculated as:

$$v_{ij} = W_j r_{ij},\ j = a, ..., m;\ i = 1, ..., n \quad (6)$$

Where:

w_j = The weight of the jth attribute or criterion

$\sum_{j=1}^{m} w = 1$

Step 4: Determine the ideal and negative-ideal solution.

$$A^* = \{v_1^*, ..., v_m^*\} = \{(\max v_{ij} / j \in I'), (\min v_{ij} / j \in I'')\} \quad (7)$$

$$A^- = \{v_1^-, ..., v_m^-\} = \{(\min v_{ij} / j \in I'), (\max v_{ij} / j \in I'')\} \quad (8)$$

Where:

I' = Associated with benefit criteria

I" = Associated with cost criteria

Step 5: Calculate the separation measures using n dimensional Euclidean distance. The separation of each alternative from the ideal solution and negative ideal solution is given as:

Table 3: Normalised data

Size	Visibility	Rich files	Scholar
0.040	0.014	0.501	0.040
0.000	0.043	1.000	0.000
1.000	1.000	0.517	1.000
0.872	0.055	0.000	0.872
0.732	0.000	0.098	0.732

$$D_i^* = \sqrt{\sum_{j=1}^{m} (v_{ij} - v_j)^2} \text{ and}$$

$$D_1^- = \sqrt{\sum_{j=1}^{m} (v_{ij} - v_j^-)^2},\ i = 1, ..., n \quad (9)$$

Step 6: Find the relative closeness to the ideal solution. The relative closeness of the alternative a_i with respect to A^* is defined as:

$$C_i^* = D_i^- / D_i^* + D_i^-),\ i = 1, ..., n \quad (10)$$

Step 7: Rank the preference order.

RESULTS AND DISCUSSION

The 4 number of criteria that should typically be considered in selecting the best university website are size, visibility, rich files and scholar. Table 1 gives number of webpages (size), number of inlinks (visibility) for university websites, rich files and scholar based on the world university highest webometrics rank in January 2012 edition. It shows that among all universities, University C had the highest universities number of webpages while Uni. J had the lowest number of webpages. The calculation of visibility requires number of inlinks to a website (Table 1). Uni. C had the highest universities visibility while Uni. Q had the lowest inlinks. Ranked first in rich files, especially from .pdf and .doc file type is Uni. N followed by Uni. I in the second place. Also, there is a tendency that the total number of pdf files exceeded the number of .doc and .ppt files.

For the scholar result, it shows that Uni. C is the leading university exceeded Uni. F in the second rank. First step of all this a approach is forming the decision matrix after that we compute h_i, d_i and w_i base on Shannon method and the result are shown in Table 4. We want to obtain a weight for each criterion by using the proposed approach. According to Eq. 1 normalized matrix data are presented. In our analysis we calculate diversity criteria and the result shows in Table 4.

The final rank of each criterion by using the entropy weighted method can be seen in Table 5. The obtained values of criterion size, visibility, rich files and scholar are 0.310458, 0.130639, 0.248445 and 0.310458, respectively. We see that the rank of size and scholar are just better

Table 4: Diversity criterion

Variables	Size	Visibility	Rich file	Scholar
	-0.6650	-0.683	-0.346	-0.6650
	-0.6930	-0.663	0	-0.6930
	0	0	-0.335	0
	-0.0887	-0.655	-0.693	-0.0887
	-0.1860	-0.693	-0.625	-0.1860
Sum	-1.6320	-2.694	-1.999	-1.6330
E(c)	0.4710	0.777	0.577	0.4710
D = E-	10.5290	0.222	0.423	0.5290

Table 5: Weight of criterion

Criteria	Weight (W) = d/total
Size	0.310
Visibility	0.131
Rich file	0.248
Scholar	0.310

Table 6: Final ranking

University	D_{max}	D_{min}	C	Ranking
A	640.225	237.647	0.271	9
B	661.823	162.880	10.198	11
C	165.022	780.427	0.825	1
D	468.150	568.357	0.548	3
E	503.998	501.821	0.499	4
F	486.380	600.383	0.552	2
H	705.554	100.135	0.124	15
I	611.179	256.054	0.295	8
J	787.378	21.293	0.026	20
K	783.092	22.824	0.028	19
L	532.571	371.568	0.411	5
M	680.569	158.974	0.189	13
N	678.537	166.711	0.197	12
O	770.179	30.450	0.038	18
P	574.565	314.246	0.354	6
Q	735.791	81.909	0.100	16
R	707.815	112.828	0.137	14
S	642.877	166.900	0.206	10
T	562.902	280.148	0.332	7
U	747.213	64.538	0.080	17

than the rank of rich file and visibility. Therefore, size locates at rank 1. Other criteria can be ranked in the same way. For problems with more complexity with a small program (for example, Excel) we can determine the rank of each criterion. In the last Table 6, the rank of each criterion can be seen.

The rich files need corresponding software for viewing. And can be downloaded from the link given alongside the links to the file itself. Also, many of the files were delivered in different formats so that the user can access the file in desired format. Then there was a university with a total of 1005600 rich files and contrastingly there was one with only 156890 files. Thus, the websites of the selected universities proved to be a mixed variety, ranging from highest to the lowest, can be found the index values is computed but before that A^*, A^-, D^* and D^- be supposed to calculate by 7-10. D^- and D^* are minimum and maximum value in table D. At this time based on the above matter the C can be accessible. C is the index value for ranking the alternatives; it can be calculated based on 10.

Now, the alternative university website are arranged in descending order according to their relative closeness values. It is observed that the Uni. C website is the best choice and followed by Uni. F based on TOPSIS method.

CONCLUSION

In this study, we have used a novel approach to measure webometrics ranking from quantitative point of view, so that, website evaluators are able to apply parameters like size, visibility, rich files and scholar for future development and webometrics to ranking problem. We used entropy and TOPSIS method to deal with weight amd ranking of the university website and presented to determine the best university website. The MCDA method which is TOPSIS based on an aggregating function representing “closeness to the ideal”. The basic principle of the TOPSIS method is that the chosen alternative should have the “shortest distance” from the ideal solution and the “farthest distance” from the “negative-ideal” solution.

IMPLEMENTATIONS

For the case studies investigated, one can observe TOPSIS viability. The TOPSIS technique is effective and simple in terms of computational implementation. Its advantages is that the algorithm does not require tuning any parameter. The TOPSIS method introduces two “reference” points but it does not consider the relative importance of the distances from these points. The TOPSIS method uses vector normalization to eliminate the units of criterion functions. While no one ranking can be accepted as definitive these webometrics ranking systems by using entropy and TOPSIS technique will remain a part of the higher education system for some time to come. It concludes by acknowledging that webometrics ranking systems are viewed differently by different stakeholders and hence can be approached in different ways.

ACKNOWLEDGEMENT

This research was supported in part by DIKTI under Grant No. 062/SP2H/DIT.LITABMAS/II/2015.

REFERENCES

Aguillo, I.F., J.L. Ortega and M. Fernandez, 2008. Webometric ranking of world universities: Introduction, methodology and future developments. Higher Educ. Eur., 33: 233-244.

Hwang, C.L. and K. Yoon, 1995. Multiple Attribute Decision Making. 1st Edn., Springer-Verlag, Berlin, ISBN: 978-0803954861.

Lai, Y.J. and C.L. Hwang, 1992. A new approach to some possibilistic linear programming problems. Fuzzy Sets Syst., 49: 121-133.

Romero, S.B. and J.C. Pomerol, 1997. [Multicriteria Decisions Theoretical Foundations and Practical Use]. University of Alcala Publisher, Alcala de Henares, Spain, ISBN:978-84-8138-180-1, (In Spainish).

Shannon, C.E., 1949. Communication theory of secrecy systems. Bell Syst. Tech. J., 28: 656-715.

Zeleny, M., 1982. Multiple Criteria Decision Making. 1st Edn. McGraw Hill, New York, USA.