International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:10, 2015

Comparative Study of Universities' Web Structure Mining

Z. Abdullah, A. R. Hamdan

***Abstract*—**This paper is meant to analyze the ranking of University of Malaysia Terengganu, UMT's website in the World Wide Web. There are only few researches have been done on comparing the ranking of universities' websites so this research will be able to determine whether the existing UMT's website is serving its purpose which is to introduce UMT to the world. The ranking is based on hub and authority values which are accordance to the structure of the website. These values are computed using two web-searching algorithms, HITS and SALSA. Three other universities' websites are used as the benchmarks which are UM, Harvard and Stanford. The result is clearly showing that more work has to be done on the existing UMT's website where important pages according to the benchmarks, do not exist in UMT's pages. The ranking of UMT's website will act as a guideline for the web-developer to develop a more efficient website.

***Keywords*—**Algorithm, ranking, website, web structure mining.

I. Introduction

IN the world we are living in today, information technology plays a very crucial role where any information desired is available just at the fingertips. With the growth of World Wide Web (WWW) technology, can be seen that existing web data covers an enormous spectrum of data from the government data, entertainment, commercial till the extend of research data. However, the vast size of data does not guarantee that users will be able to acquire wanted data easily. In reality, there is an opinion [1] saying that 99% data on the WWW is useless to 99% of web browsers. In a way, by looking at the various number of websites, this does prove that somewhere in the tangled mane of websites, exist some precious useful information but this information is impossible to be recognized by users solely by depending on intuition so this is where web mining comes into picture.

Web mining is defined as a converging research area from several research communities, such as database, information retrieval, machine learning and natural language processing [2]-[5]. Web mining is the use of data mining techniques to automatically dis-cover and extract information from Web documents and services. The objects of web mining are web resources, which can be web documents, web log files, structure of web sites and structure of web documents themselves. Web mining can be categorized into three areas of interest based on which part of the web to mine. These categories [3]-[5] are Web usage mining, Web content mining and Web structure mining.

This research focuses on Web structure mining which can be regarded as the process of discovering structure information from the Web [3]-[5]. This type of mining can be further divided into two kinds based on the kind of structural data used. The first type is pattern extraction using hyperlinks. A Hyperlink is a structural unit that connects a Web page to different location, either within the same Web page or to a different Web page. A hyperlink that connects to a different part of the same page is called an Intra-Document Hyperlink, and a hyperlink that connects two different pages is called an Inter-Document Hyperlink. The second type is the Document Structure where the content within a web page can also be organized in a tree-structured format, based on the various HTML and XML tags within the page.

This research will focus on the application of Web structure mining on Universiti Malaysia Terengganu's (UMT) website. The authority and hub value of the website will be determined using web-searching algorithms which are HITS and SALSA.

For HITS, Kleinberg [6] distinguished between two types of Web pages which pertain to a certain topic which are hub and authority. Hubs are primarily resource lists, linking to many authorities on the topic possibly without directly containing the authoritative information. According to this model, hubs and authorities exhibit a mutually reinforcing relationship: good hubs point to many good authorities, and good authorities are pointed at by many good hubs. In light of the mutually reinforcing relationship, hubs and authorities should form communities, which can be pictured as dense bipartite portions of the Web, where the hubs link densely to the authorities.

SALSA is a stochastic approach developed by Lempel and Moran [7] in which the coupling between hubs and authorities is less tight. The intuition behind our approach is the following: consider a bipartite graph G, whose two parts correspond to hubs and authorities, where an edge between hub r and authority s means that there is an informative link from r to s. Then, authorities and hubs pertaining to the dominant topic of the pages in G should be highly visible (reachable) from many pages in G. Thus, we will attempt to identify these pages by examining certain random walks in G, under the proviso that such random walks will tend to visit these highly visible pages more frequently than other, less connected pages.

This research is very important because most of the website nowadays has been developed all under the will and idea of the web developer oneself without any references to how other

Z. Abdullah is with the University of Malaysia Terengganu, 21030, Kuala Terengganu, Malaysia (corresponding author. phone: 603-6683536; fax: 603-669-4100; e-mail: zailania@ umt.edu.my).

A. R. Hamdan is with the University of Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia (e-mail: arh@ftsm.ukm.my).

developers developing website of the same category with higher ranking had done theirs. This is just the same scenario with UMT's website, where by comparing the search result on the WWW, UMT is ranked 85th while UM is ranked 5th on the search result list. By looking at the resulted ranking, it is obvious that UMT's website could do with this research to mend the ranking of the university's website thus achieving the notion of introducing more of UMT to the world.

Up to this date, researches were done to compare the used algorithms and the means to improve the algorithms; none were done to compare the resulted reading of the algorithms [8]-[12]. Another problem is that existing researches were done randomly on data extracted from the WWW, none of the data used were specifically focusing on evaluating the structure of any university's website [8]-[15].

Aim of this research is to understand the web structure mining algorithms which are HITS and SALSA to enable the development of such algorithms using software MATLAB 7.0. The algorithms are the tools to determine the hub and authority values of studied websites. The readings produced by the algorithms will drive the research where they will be used to evaluate the web structure on UMT's website through comparison with three websites which are Universiti Malaya (UM), Harvard and Stanford. The scope of this research is to use UMT as the research case and three other universities' website (UM, Harvard and Stanford) as benchmarks. As well to develop HITS and SALSA's algorithms using MATLAB 7.0.

The remaining part of the paper is organized as follows. Sections II present the related works. Section III elaborates the methodology. Section IV reports the result and discussion.

II. Related Works

For HITS and SALSA, the obvious comparison is the effectiveness of SALSA in coming up with more useful information for a query to be compared to HITS [5],[8]. It was shown that in searching for a single query, the Mutual Reinforcement approach of HITS ranked many irrelevant pages as authorities whereas SALSA which is less vulnerable to the Mutual Reinforcement Approach produced better finding. However, since the implementation were based on query, this aspect of the algorithms will only be used as a guidelines since this research will be based on exactly universities' web pages.

Several researches have even introduced some modification to the existing algorithms [10], [11], [14], [15] where in all the research, the modified version was proven to be producing more detailed results. The modified versions were intended to be implement in favor of the original algorithms but still no such scenarios are seen in any part of the WWW since the original algorithms are still the only applied version of the algorithms, so for this research the modified version of both algorithms are totally not being taken into account.

Application of HITS and SALSA are apparent on the WWW, where examples of websites that use these applications are www.linkexchange.com, www.amazon.com, www.amazoncity.com and www.amazonrecords.com. These websites all provide sales and online purchasing services only. It is yet to be found an application of HITS and SALSA on university's website and that is what this research will try to initiate.

Some researches were done using data based on large uncategorized group of websites [1], [5], [8], [10], [13], [14] where the studied websites are randomly selected from a large web graph which was induced by millions of web crawlers. This kind of researches bear no result regarding the web structure of the websites where they only produce the readings of the hub and authority without explanation of what elements those contribute to the high hub and authority values. To combat the lack of explanation on the responsible elements that produce high hub and authority value, this research will present the explanation in form of categories of the universities web pages.

This research is important because in the end, the result will show the structure on UMT's website and the comparison between the rankings of the websites in determining on how UMT's website can be improve in term on increasing the hub and the authority value on the WWW so the website will serve the real function it was built for, to introduce UMT to the world.

III. Methodology

In developing the HITS (Fig. 1) and SALSA (Fig. 2) algorithms, software MATLAB 7.0 is the most appropriate to be used since it provides simultaneously application of programming and mathematics. It also allows the user to conjure up images of the results in all intended forms matrix, graph, images, simulation and etc. For this research, the algorithms, the mapping of web structure matrices of the universities' websites, and the resulted hub and authority values were executed using this software.

Hit Authority

$$X_i^{(k)} = \sum_{i.e_{i,j} \in E} y_j^{(k-1)}$$

Hit Hub

$$X_i^{(k)} = \sum_{i.e_{i,j} \in E} y_j^{(k-1)}$$

i = page

X_i = authority score

Y_i = hub score

$e_{i,j}$ = directed edge from node i to node j

E =set of all directed edges in the web graph

In matrix form

$$L_{i,j} - \begin{cases} 1, \text{if there exists an edge from node i to node j} \\ 0, \text{if there non exists an edge from node i to node j} \end{cases}$$

The initial equation can be simplified by substitution to

Authority matrix

$$x^{(k)} = L^T y^{(k-1)} \rightarrow x^{(k)} = L^{(T)} L x^{(k-1)}$$

Hub matrix

$$y^{(k)} = L x^{(k)} \rightarrow y^{(k)} = L L^{(T)} L y^{(k-1)}$$

Fig. 1 HITS Formula

Fig. 1 is the formula of HITS. The hub and authority values are obtained by multiplication of the web structure matrix, L

with its adjacency matrix, LT in two different sequences. For hub it is LLT and for authority it is LTL. All hub and authority scores are checked with the eigenvector before obtaining the final value to ensure the final sum is equal to one (1) which if not, it shows that the web structure matrix used as L is incomplete or contain error.

$$h_{i,j} = \sum_{\left\{k|(i_h,k_a),(j_h,k_a)\in G \frac{1}{\deg(i_h)}\bullet\frac{1}{\deg(k_a)}\right\}}$$

(a)

$$a_{i,j} = \sum_{\left\{k|(k_{h,i_a}),(k_h,j_a)\in G \frac{1}{\deg(i_a)}\bullet\frac{1}{\deg(k_h)}\right\}}$$

(b)

Fig. 2 SALSA Formula (a) SALSA hub (b) SALSA authority

Fig. 2 is the formula of SALSA. Both equations explain about a certain page *k* points to both pages *i* and *j,* hence page *j* is reachable from page *i* by two steps which are retracting along the link from page *k* till page *i* and then following the link from page *k* till page *j*. Hub value will be the value of nonzero entry of the matrix (G) divided by the sum of entries in row of G whilst the authority value will be the value of nonzero element of G divided by sum of entries in column of G.

The methodology used for this research is build of five-level steps. The steps are:

A. Problem Definition

The main problem of this research is to acquire the hub and authority values of the studied universities websites in order to determine which category of pages that contributes to the high hub and authority values.

B. Data Selection

For this research data was collected from four websites where UMT as the studied website and the other comparison websites, UM, Harvard and Stanford.

C. Preprocessing of Data

URL for each page of selected websites was manually extracted from WWW. The raw data then was manually transformed into binary matrix form (0 represent no hyperlink and 1 represent existing hyperlinks between mapped pages) (Fig. 3). The web structure mining algorithms, HITS and SALSA were developed using MATLAB.

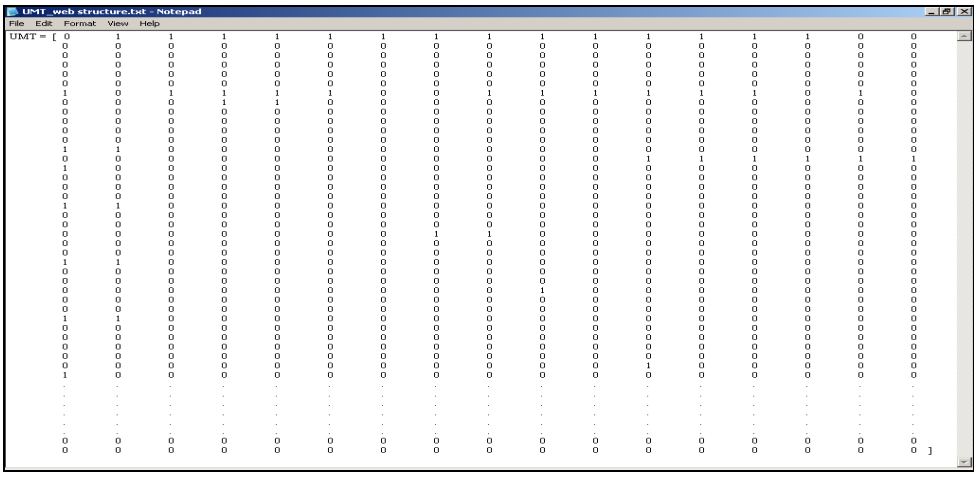
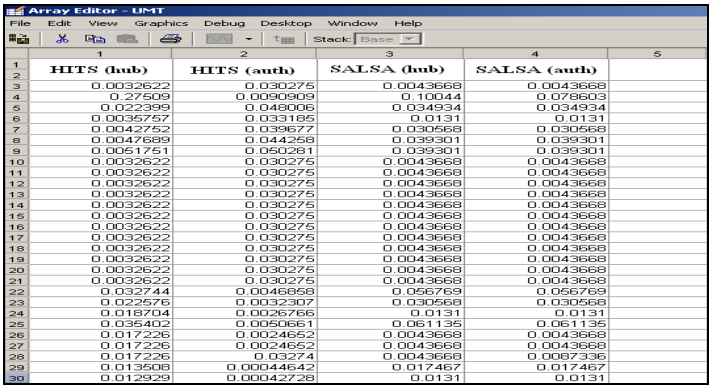


Fig. 3 Web Structure Matrix of UMT's Website



	1	2	3	4	5
1	HITS (hub)	HITS (auth)	SALSA (hub)	SALSA (auth)	
2	0.0032622	0.030275	0.0043668	0.0043668	
3	0.27609	0.0090909	0.10044	0.078603	
4	0.022399	0.048006	0.034934	0.034934	
5	0.0035757	0.033185	0.0131	0.0131	
6	0.0042752	0.039677	0.030568	0.030568	
7	0.0047689	0.044258	0.039301	0.039301	
8	0.0051751	0.050281	0.039301	0.039301	
9	0.0032622	0.030275	0.0043668	0.0043668	
10	0.0032622	0.030275	0.0043668	0.0043668	
11	0.0032622	0.030275	0.0043668	0.0043668	
12	0.0032622	0.030275	0.0043668	0.0043668	
13	0.0032622	0.030275	0.0043668	0.0043668	
14	0.0032622	0.030275	0.0043668	0.0043668	
15	0.0032622	0.030275	0.0043668	0.0043668	
16	0.0032622	0.030275	0.0043668	0.0043668	
17	0.0032622	0.030275	0.0043668	0.0043668	
18	0.0032622	0.030275	0.0043668	0.0043668	
19	0.0032622	0.030275	0.0043668	0.0043668	
20	0.0032622	0.030275	0.0043668	0.0043668	
21	0.0032622	0.030275	0.0043668	0.0043668	
22	0.032744	0.0046858	0.056769	0.056769	
23	0.022576	0.0032307	0.030568	0.030568	
24	0.018704	0.0026766	0.0131	0.0131	
25	0.035402	0.0050651	0.061135	0.061135	
26	0.017226	0.0024652	0.0043668	0.0043668	
27	0.017226	0.0024652	0.0043668	0.0043668	
28	0.017226	0.03274	0.0043668	0.0087336	
29	0.013508	0.00044642	0.017467	0.017467	
30	0.012929	0.00042728	0.0131	0.0131	

Fig. 4 Example of Hub and Authority value of UMT's Website

D. Application of Data

The complete mapped matrices were executed using the two algorithms, HITS and SALSA. The readings of the algorithms (Fig. 4) are in the form of matrix (raw data which are still not meaningful until being processed). The readings are values of hub and authority for each website.

E. Processing of Data

The raw data from the reading are turned into meaningful data by ranking the top 15 pages of each website. Graphs are produced to show the differences between case study and the comparison websites in the top 15 pages ranking according to algorithms. The comparisons are recorded as the result of the research. The framework used for this research is as illustrated in Fig. 5.

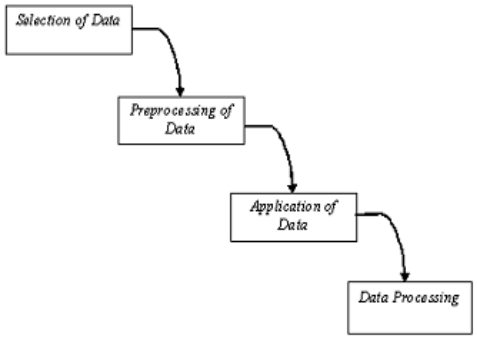


Fig. 5 Overview Model of the Framework

IV. Result and Discussion

The experiments were carried out on Intel® CoreTMi5-321M CPU at 2.50 GHz speed with 4GB RAM, running on Window 7 Home Premium. All algorithms have been developed using Java as a programming language and NetBeans IDE 8.0 with JDK 1.7.09 as platform.

Table I and Fig. 6 shows that for UMT's website, in term of value, the highest ranked pages fall under the 'General' category (HITS authority, 0.7916) whilst in term of highest number of pages, most pages in top 100 pages are under 'Academics' category (46 pages).

TABLE I
Top 100 Ranked Pages of UMT's Website

Web page Category	Code	No of Pages	Weight Value of Algorithms			
			HITS (hub)	HITS (authority)	SALSA (hub)	SALSA (authority)
Administration	1	18	0.1030	0.1065	0.0974	0.1023
General	2	22	0.7688	0.7916	0.2825	0.2955
Academics	3	46	0.1297	0.1019	0.1331	0.1591
News & Events	4	8	0.0067	0.0002	0.0721	0.0721
Facilities	5	-	-	-	-	-
Athletics	6	-	-	-	-	-
Arts	7	-	-	-	-	-
Others	9	6	0.0067	0.0001	0.0541	0.0541
Total		100	1.0149	1.0003	0.6392	0.6831

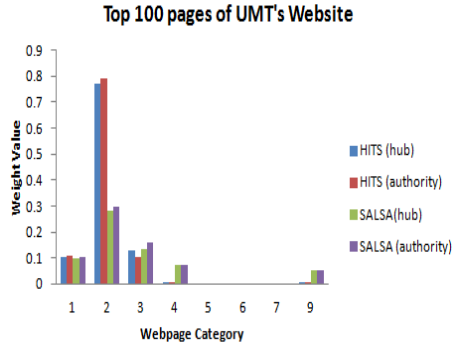


Fig. 6 Top 100 Ranked Pages of U MT's Website

Table II and Fig. 7 shows for UM's website, in term of value, the highest ranked pages fall under the 'General' category (HITS authority, 0.6556) whilst in term of highest number of pages, most pages in top 100 pages are under 'Academics' category (35 pages).

TABLE II
Top 100 Ranked Pages of UM's Website

Web page Category	Code	No of Pages	Weight Value of Algorithms			
			HITS (hub)	HITS (authority)	SALSA (hub)	SALSA (authority)
Administration	1	8	0.1818	0.0053	0.1528	0.1528
General	2	31	0.3616	0.6556	0.3581	0.3362
Academics	3	35	0.3020	0.1156	0.3406	0.3450
News & Events	4	8	0.1271	0.0021	0.0349	0.0349
Facilities	5	13	0.0236	0.0507	0.0568	0.0568
Athletics	6	-	-	-	-	-
Arts	7	2	0.0005	0.0049	0.0007	0.0007
Alumni	8	-	-	-	-	-
Others	9	3	0.0008	0.0074	0.0003	0.0003
Total		100	0.9974	0.8416	0.9442	0.9267

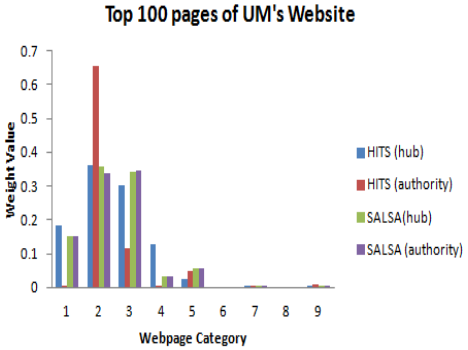


Fig. 7 Top 100 Ranked Pages of UM's Website

Table III and Fig. 8 shows that in term of value, the highest ranked pages fall under the 'General' category (SALSA hub, 0.7705) whilst in term of highest number of pages, most pages in top 100 pages are under 'Administration' category (33 pages).

TABLE III
TOP 100 RANKED PAGES OF HARVARD'S WEBSITE

Web page Category	Code	No of Pages	Weight Value of Algorithms			
			HITS (hub)	HITS (authority)	SALSA (hub)	SALSA (authority)
Administration	1	33	0.1002	0.0911	0.0450	0.0451
General	2	22	0.6667	0.5299	0.7705	0.7479
Academics	3	24	0.1204	0.1031	0.0779	0.0779
News &Events	4	-	-	-	-	-
Facilities	5	8	0.0228	0.0228	0.0164	0.0164
Athletics	6	-	-	-	-	-
Arts	7	13	0.0383	0.0670	0.0471	0.0471
Alumni	8	-	-	-	-	-
Others	9	-	-	-	-	-
Total		100	0.9484	0.8139	0.9569	0.9344

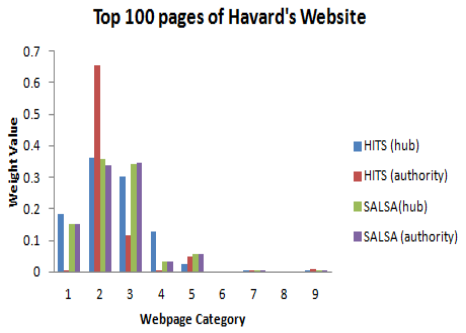


Fig. 8 Top 100 Ranked Pages of Harvard's Website

Table IV and Fig. 9 shows that in term of value, the highest ranked pages fall under the 'Academics' category (SALSA hub and authority, both 0.3924) whilst in term of highest number of pages, most pages in top 100 pages are under 'Academics' category (35 pages).

Table V and Fig. 10 are summary of the top 100 ranked pages in term of weight value accordingly by algorithms:

1. For UMT's website, both HITS and SALSA for hubs and authority values assigned pages under 'General' category with the highest weight.
2. For UM's website, HITS for hubs and authority values also SALSA for hub value assigned the highest weight to pages under 'General' category whilst SALSA for authority value assigned the highest weight to pages under 'Academics' category.
3. For Harvard's website, all HITS and SALSA values for hubs and authority are assigned to pages under 'General' category.
4. For Stanford's website, all values of HITS and SALSA, both hubs and authority are assigned to pages under 'Academics' category.

TABLE IV
TOP 100 RANKED PAGES OF STANFORD'S WEBSITE

Web page Category	Code	No of Pages	Weight Value of Algorithms			
			HITS (hub)	HITS (authority)	SALSA (hub)	SALSA (authority)
Administration	1	12	0.0963	0.0963	0.0965	0.0965
General	2	27	0.3246	0.3246	0.3199	0.3199
Academics	3	35	0.3906	0.3906	0.3924	0.3924
News & Events	4	2	0.0249	0.0249	0.0256	0.0256
Facilities	5	4	0.0480	0.0480	0.0511	0.0511
Athletics	6	6	0.0746	0.0746	0.0767	0.0767
Arts	7	3	0.0373	0.0373	0.0383	0.0383
Alumni	8	7	0.0011	0.0011	0.0011	0.0011
Others	9	4	0.0006	0.0006	0.0006	0.0006
Total		100	0.998	0.998	1.0022	1.0022

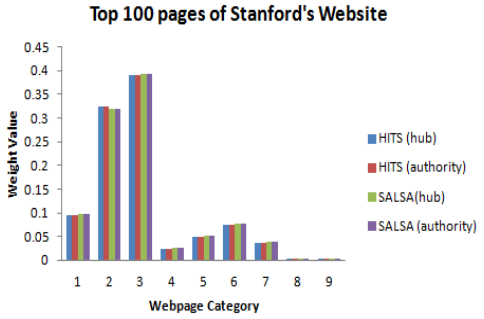


Fig. 9 Top 100 ranked pages of Stanford's Website

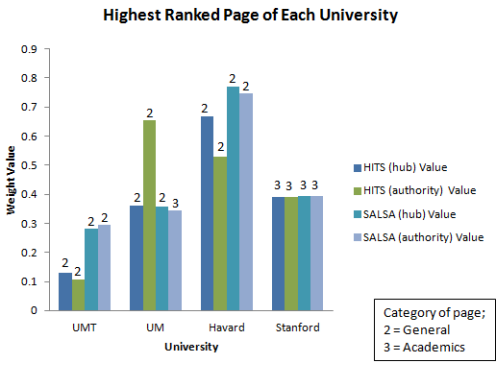


Fig. 10 Highest Ranked Pages of Each University

TABLE V
HIGHEST RANKED PAGES *OF* EACH UNIVERSITY

University	HITS (hub)		HITS (authority)		SALSA (hub)		SALSA (authority)	
	Value	Category	Value	Category	Value	Category	Value	Category
UMT	0.1297	2	0.1065	2	0.2825	2	0.2955	3
UM	0.3616	2	0.6556	2	0.3581	2	0.3450	3
Havard	0.6667	2	0.5299	2	0.7705	2	0.7479	3
Stanford	0.3906	2	0.3906	3	0.3924	3	0.3924	3

Category; 2: General, 3: Academics

Table VI shows the pages under which category contribute most to the top 100 ranked pages of each university. For UMT it is 'Academics' (46 pages), UM is also 'Academics' (35 pages), 'Harvard' is Administration (33 pages) and Stanford is 'Academics' (35 pages).

TABLE VI
TOP 100 RANKED PAGES ACCORDING TO 'NO. OF' PAGE

Web page Category	Code	University	No. of Pages
Administration	1	UMT	18
		UM	8
		Harvard	33
		Stanford	12
General	2	UMT	22
		UM	31
		Harvard	22
		Stanford	27
Academics	3	UMT	46
		UM	35
		Harvard	24
		Stanford	35
News & Events	4	UMT	8
		UM	8
		Harvard	-
		Stanford	2
Facilities	5	UMT	-
		UM	13
		Harvard	8
		Stanford	4
Athletics	6	UMT	-
		UM	-
		Harvard	-
		Stanford	6
Arts	7	UMT	-
		UM	2
		Harvard	13
		Stanford	3
Alumni	8	UMT	-
		UM	-
		Harvard	-
		Stanford	7
Others	9	UMT	6
		UM	3
		Harvard	-
		Stanford	4

The major difference produced by the comparison according to Table VI, is that the three comparison websites have listed pages under two other category of pages in their top 100 raked pages which is lacking in UMT's website. The categories are:

1. Facilities (UM: 13 pages, Harvard: 8 pages, Stanford: 4 pages)
2. Arts(UM: 2 pages, Harvard: 13 pages, Stanford: 3 pages)

The comparison shows that UMT's website still needs to be 'touch-up' and among the manners that can be implement are:

First, websites that contain information which can be regarded as 'additional information' such as 'Facilities' and 'Arts' need to be emphasis because it is clear that these categories contribute quite a portion of the hub and authority value as apparent in the three comparison websites. Up till the time of this research is conducted, UMT's website has no specific pages under these two categories yet.

Second, for the highest ranked pages according to weight values, even though UMT's website has assigned pages under 'General' category with the highest values which is the same case for UM's and Harvard's website, it is apparent that the values produced by UMT's website by far are the lowest (Example: HITS hub values; UMT=0.1297, UM=0.3616, Harvard=0.6667). This shows that even though the emphasis has already been given to the right category, there are still elements missing (such as hyperlink to other meaningful pages or meaningful information on the specific website itself) which resulted in low weight values.

Third, UMT's website can also create a direct link from page 'home' to other pages which can increase the hub value like what has been done by Harvard and Stanford. The main pages of each category is linked to every sub-page in that exact category and also linked to sub-pages of other categories.

Fourth, as in case of Stanford's website, the weight values for all categories are very consistent and the website listed pages of category 'Athletics' (HITS: 0.0746, SALSA: 0.0767) and 'Alumni' (all HITS and SALSA: 0.0011) under the top 100 ranked pages of its website. Even though the other two comparison website have no pages under these two categories in their top 100 ranked pages, UMT's website can still focus on these two categories since without doubt, Stanford is among the best well-known university in the world and they have given emphasis on this.

V. CONCLUSION

So far, there are no research have been done focusing on web structure mining of universities' websites in comparing what are the essential pages to ensure the value of hub and authority are high. This research is done by focusing on four universities' websites, UMT as the case study and three other websites, UM, Harvard and Stanford as the comparison to identify the important pages a university's website should emphasize on. Two web structure algorithms, HITS and SALSA have been developed. The methodology used was simple steps where the data structure of each website is manually mapped into matrix to enable the algorithms to run through the matrix and produce the top-ranked pages of each university. The result shown that UMT's website is absolutely lacking in two main categories of pages which was emphasized in the comparison websites which are 'Facilities' and 'Arts' also two minor categories which are 'Athletics' and 'Alumni'. This finding will enable the UMT's website developer to improve the structure of the website and create the important non-existing pages as what has been done by the comparison universities.

ACKNOWLEDGMENT

This work is supported by the research grant from Research

Acceleration Center Excellence (RACE) of Universiti Kebangsaan Malaysia.

REFERENCES

[1] A. Arasu, J. Cho, H. Garcia-Molina, A. Paepcke, and S. Raghavan, "Searching the Web," *ACM Transactions on Internet Technology (TOIT)*, 1 (1), pp. 2-43, 2001.

[2] B.J. Jansen, A. Spink, C. Blakely, and S. Koshman, "Defining a Session on Web Search Engines," *Journal of the American Society for Information Science and Technology*, 58(6), pp. 862–871, 2007.

[3] J. Srivastava, P. Desikan, and V. Kumar, "Web Mining-Accomplishments & Future Directions," University of Minnesota. 2000

[4] J. Fürnkranz, "Web Mining," *Data Mining and Knowledge Discovery Handbook*, pp. 913-930, Springer-Verlag, 2010.

[5] M. Eirinaki, "Web Mining: A Roadmap," *Technical Report*, DB-NET 2004, at http://www.engr.sjsu.edu/meirinaki/papers/NEMIS.pdf

[6] J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *Proceeding of the 9th ACM SIAM Symposium on Discrete Algorithms*, pp. 668–677, 1998.

[7] M. Lan, "Algorithms and Applications of Preference Based Ranking for Information Retrieval," *Ph.D Thesis*, 2005.

[8] M. Najork, "Comparing the Effectiveness of HITS and SALSA," *Proceeding of 16th ACM Conference on Information and Knowledge Management* (CIKM), 2007.

[9] R. Lempel, and S. Moran, "Rank-Stability and Rank-Similarity of Link-Based Web Ranking Algorithms in Authority-Connected Graphs," *Information Retrieval*, pp. 245-264, 2005.

[10] Y Duan, J Wang, M Kam, J Canny, "Privacy preserving link analysis on dynamic weighted graph," *Computational & Mathematical Organization Theory*, 11 (2), 141-159, 2005.

[11] Z. Chen, L. Tao, J. Wang, L. Wenyin, and W. Ma, "A Unified Framework for Web Link Analysis," Proc. 3rd International Conference on Web Information Systems Engineering (WISE2002), Singapore (regular paper), pp. 63-72, Dec 2002.

[12] A. Borodin, G. O. Roberts, J. S. Rosenthal, and P. Tsaparas, "Finding Authorities and Hubs from Link Structures on the World WideWeb," *Proceedings of the 10th International World Wide Web Conference*, pp. 415-429, 2001.

[13] A.N. Langville, and C.D. Meyer, "A Survey of Eigenvector Methods for Web Information Retrieval," Journal SIAM review, 47(1), 135-161, 2005.

[14] A. Farahat, T. LoFaro, J.C. Miller, G. Rae, L.A. Ward, "Authority rankings from HITS, PageRank, and SALSA: Existence, uniqueness, and effect of initialization," *SIAM Journal on Scientific Computing*, 27 (4), 1181-1201, 2006.

[15] J.C. Miller, G. Rae, and F. Schaefer, "Modifications of Kleinberg's HITS Algorithm Using Matrix Exponentiation and Web Log Records," *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 444-454, 2001.