# University website quality comparison by using non-parametric statistical test: a case study from Malaysia

P.D.D. Dominic* and Handaru Jati

Department of Computer and Information Science,
Universiti Teknologi PETRONAS,
31750 Tronoh, Perak, Malaysia
E-mail: pdddominic@yahoo.com
E-mail: dhanapal_d@petronas.com.my
E-mail: handaruj@yahoo.com
*Corresponding author

Suhaiza Hanim

School of management,
University Sains Malaysia,
11800 Penang, Malaysia
E-mail: shmz@usm.my

**Abstract:** This research conducts some tests to measure the quality of Malaysian University website via web diagnostic tools online. We propose a methodology for determining and evaluating the best university website based on many criteria of website quality, consist of linear weightage model (LWM), analytical hierarchy process, fuzzy analytical hierarchy process (FAHP) and new hybrid model (NHM). This NHM has been implemented using combination of LWM and FAHP to generate the weights for the criteria which are better and more fairly preference. The result of this study confirmed that most of the Malaysian University websites are neglecting performance and quality criteria. By applying hybrid model between LWM and FAHP approach for website evaluation has resulted in significant acceleration of implementation, raised the overall effectiveness with respect to the underlying methodology and ultimately enabled more efficient and significantly equal or better procedure compared with other methods.

**Keywords:** performance; university website; quality; web diagnostic; hybrid model.

**Reference** to this paper should be made as follows: Dominic, P.D.D. Jati, H. and Hanim, S. (2013) 'University website quality comparison by using non-parametric statistical test: a case study from Malaysia', *Int. J. Operational Research*, Vol. 16, No. 3, pp.349–374.

**Biographical notes:** P.D.D. Dominic obtained his MSc degree in Operations Research in 1985, MBA from Regional Engineering College, Tiruchirappalli, India in 1991, Post-graduate Diploma in Operations Research in 2000 and completed his PhD in 2004 in the area of Job Shop Scheduling at Alagappa University, Karaikudi, India. Since 1992, he has held the post of Lecturer in the Department of Management Studies, National Institute of Technology

(formally Regional Engineering College), Tiruchirappalli, India. Currently, he is working as an Associate Professor in the Department of Computer and Information Science, University Teknologi PETRONAS, Malaysia. His fields of interest are operations management, KM and decisions support systems. He has published more than 50 technical papers in international, national journals and conferences.

Handaru Jati obtained his MSc degree in Information Technology in 2005 and MBA degree in 2001 from Gadjah Mada University, Indonesia. Currently, he is pursuing his Doctoral degree from Universiti Teknologi PETRONAS in Department of Computer and Information Science, Malaysia. Since 1999, he has held the post of Lecturer in the Department of Electronics and Informatics Engineering at Yogyakarta State University, Indonesia (formally Yogyakarta Vocational Education Institute). His fields of interest are software quality, image processing and e-learning evaluation. He has published technical papers in international and national conferences.

Suhaiza Hanim obtained her MSc degree in Management Science specialised in Operations Research in 1995 and completed her PhD in 1998 in the area of Operations Management from Lancaster University, UK. Currently, she is working as an Associate Professor in the School of Management, Universiti Sains Malaysia, Malaysia. She got more than 15 years of teaching and research experience. Her fields of interest are operations management, decisions support systems and supply chain management. She is in the editorial and reviewer board for more than 10 international journals including *IJOSCM*, *IJVCM*. She has published more than 100 technical papers in international, national journals and conferences.

## 1 Introduction

The question of website quality has been defined by many disciplines in three distinct ways: the information value of the content provided (library and information science), the design of such a site (information systems and technology, media studies) and the usability of the interface (mediated communication). Each definition of quality leads to lists of criteria about what constitute a quality site. All of these criteria from multiple studies on web quality form a comprehensive tool for evaluating the quality of a website that would serve to assess its trustworthiness (McInerney, 2000). There is a principle that 'if information can pass a test of quality, it is most likely to prove trustworthy' and because of this belief, higher quality website should have higher credibility. Thus, the challenge is how to create a method that will guide the internet user to evaluate a website without needing a lot of time. The method needs a lot of time and cautious consideration. It takes more than 1 hr to examine a website thoroughly and apply the criteria of quality. This time dedication may be available to information professionals, but public users may not be willing to spend the same amount of time.

The evaluation of a website in terms of quality lacks a single-point definition. It is the combination of various factors: aesthetic, logic, technology and many other factors. There are many scopes of quality and each measure will pertain to a particular website in varying degrees. Here are some of them: the first factor is time; credible site should be updated frequently. The information about the latest update should be included on the homepage. If the information in the website is not updated frequently, the visitor could simply assume that perhaps the site manager has no time to update the site.

The second factor is structural; all website components should hold together and all website internal and external links should work well. Clear navigation, legible content, clean page layouts, simple instructions and easy search functions are factors that contribute to user friendliness of website. Broken links on the webpage are also another factor that can downgrade the website quality. Each page usually has references or links or connections to other pages and these links connect to the internal or external websites. Users expect each link to be valid, meaning that it leads successfully to the intended page or another resource. In the year of 2003, it was discovered that about 1 out of every 200 links disappeared each week from the internet (McCowen et al., 2005).

The third factor is content or search engine friendliness; the number of the links or link popularity is one of the off-page factors that search engines are looking for to determine the value of the webpage. Internet users are increasing all over the world and online businesses are absolutely on an immense raise. To generate income through merchandising sales, institutions need to have quality web traffic first. Search engines are important to websites success. At the very least, website should be search engine friendly. Search engines should be able to easily extract the content available for public and display the relevant pages to fulfil the search queries. Major search engines have their own way of defining relevant search results for particular key phrase. Approximately 85% of all traffics to websites and 70% of all online business transactions originate from a search engine or directory (such as Google, Yahoo or MSN). Quality web traffics can be obtained by improving website ranking on most of the search engines. To improve website ranking, there are some steps to do, Firstly, by analysing who are target audiences and then analysing what are the keywords or phrases the target audiences are using while searching the web. If the website has a very rich content with those keywords, there will be a higher chance to improve the website ranking on any search engine. Secondly, by updating the content regularly and this action can improve the website ranking remarkably. This is because most of the search engine algorithms give top ranking while indexing the websites if the content is updated frequently. Search engines require a website to have at least two links pointing to the site before they will place it to their index. The idea is that when a website has increased its link popularity, then it has also proved that the website has high quality content. Number of links to website improves access growth and helps to generate traffic (Page et al., 1998).

The fourth factor is response time; a website server should respond to a browser request within certain parameters. Popular sites averaged 52 objects per page, 8.1 of which were ads, served from 5.7 servers (Krishnamurthy and Wills, 2006) and object overhead now dominates the latency of most web pages (Yuan et al., 2005). Following the recommendation of the hyper text transfer protocol (HTTP) 1.1 specification, browsers typically default to two simultaneous threads per hostname. As the number of HTTP requests required by a web page increase from 3 to 23, the actual download time of objects as a percentage of total page download time drops from 50% to only 14%. Table 1 shows the ranking of Malaysian Universities based on a report published by webometric and the ranking position is Universiti Sains Malaysia (USM), Universiti Kebangsaan Malaysia (UKM), Universiti Putra Malaysia (UPM), Universiti Utara Malaysia (UUM) and last Universiti Teknologi PETRONAS (UTP).

**Table 1** Ranking of the Malaysian Universities website based on webometric survey

| *University* | *Address* | *Ranking* |
|---|---|---|
| USM | www.usm.my | 675 |
| UKM | www.ukm.my | 963 |
| UPM | www.upm.edu.my | 979 |
| UUM | www.uum.edu.my | 1,329 |
| UTP | www.utp.edu.my | 2,515 |

The fifth factor is stickiness, which is the ability to ensure that the internet user sticks on the website page for a longer period of time. A sticky website is a place that people will come to visit again. By having repeat visitors, this strategy can increase exposure to product or service hence creates more sales. The positive impacts to have a sticky website are: repeat traffic impact on increased sales, create one-to-one relationships and develop performance through feedback.

The sixth factor is design, a site does not only need to make sense visually, it should also appear the same on all web browsers (such as Internet Explorer, Opera and Firefox) and across all computer platforms (PC and Mac). Good design should make a site easy to use and an effective site design should communicate a brand and help to accomplish the site's objectives and goals. However, creating website with a good design is subjective and it is only through repetitive efforts and testing that we can figure out what works best for the intended audience.

The last factor is performance. Technology continues to make important impact in service industries and fundamentally shapes how services are delivered (Durkin, 2007). There are also many factors influence the performance of the web and most of them are outside the control of website designer. Download time of a website is determined by web page design, web server, hardware of the client, software configuration and characteristics of the internet router which connects user and the website. One research finding mentioned that a website which has slow download time is less attractive compared to a website with faster download time (Ramsay et al., 1998). Currently, the average connection speed is 5 kbps (kilobytes per sec) and this gives an implication that one web page with 40 kb page size will be downloaded within 8 sec. This matter is in accordance with the '8-sec rule', that 8-sec period is a normal time for loading a webpage and it is not be tolerable by the user. This fact is supported by many research results mentioning that the mean of tolerable download time by the user side is 8.57 sec with standard deviation of 5.9 sec (Bouch et al., 2000). This also shows that providing information related with waiting time is very important for the users. Therefore, for long download time, it is better to provide information about how many per cent of the webpage already downloaded and how much time needed to complete this task. Another important aspect is information fit-to-task, which means that information presented on a website is accurate and appropriate for the task at hand (Loiacono et al., 2007). Good architecture is fundamental to deal with a website's requirements, to ensure structural scalability, flexibility, security and to fulfil performance demands currently and in the future. A completed site should comply with acknowledged programming standards. As the web keeps on growing as a competitive tool for business applications, there is a need to comprehend the relationship between business performance and web usability. Most of the previous researches have discussed the website development from a set of usability factors (Green and Pearson, 2006; Seffah et al., 2006). Online accessibility test can be used to examine whether the web portals have accessibility errors on their respective web pages and the world wide web consortium (W3C) rules are divided into three priority levels which will influence the level of website accessibility. If a website cannot satisfy the second priority, then users will have some problems to access the website; however, if the website already satisfied the third criteria, then users will have a little difficulty to access the web (Loiacono and McCoy, 2004).

The problem of a decision-maker consists of evaluating a set of alternatives to find the best one, to rank them from the best to the worst and to describe how well each alternative meets all the criteria simultaneously. There are many methods for determining the ranking of a set of alternatives in terms of a set of decision criteria. In a multi-criteria approach, the analyst seeks to build several criteria using several points of view. Multiple-criteria decision-making (MCDM) is one of the most widely used decision methodologies in science, business and governments. In the problem of website selection, the decision-maker has a large set of criteria for selecting websites. The problem is to compare the various criteria and to determine their relative importance through pair-wise comparison between each pair of them, examples for the application of the MCDM were used to solve the problem of portfolio selection in Istanbul stock exchange (Tiryaki and Ahlatcioglu, 2009) to integrate an active set algorithm optimisation for portfolio selection into a multi-objective evolutionary algorithm (Branke et al., 2009) and to create portfolio selection as three objective optimisation problem to find trade-offs between risk, return and the number of securities in the portfolio (Anagnostopoulos and Mamanis, 2010).

## 2 Literature review

### *2.1 Website evaluation studies*

The website evaluation can be approached from users, website designer/administrator or both together (Sayar and Wolfe, 2007). From the user's perspective, most of the studies on website evaluation focus on the factors for successful websites. These researches concentrate on the development of a website evaluation tool. These studies search for design and content elements of a successful website using the exploratory study. The main areas for the website quality evaluation are: function, usability, efficiency and reliability (Olsina et al., 2001). Website quality evaluation method is used to test six university sites from different countries (Olsina et al., 2001). Website architecture is classified into content and design (Huizingh, 2000) and each classification is specified into evaluation criteria according to the characteristics and perception of a website.

From the website designer or administrator's perspective, the website evaluation focuses on the web usability and accessibility. The website evaluation model is based on the study of the user-centred development and evaluation approach. This study attempts to develop the methodology and tool for the website quality evaluation from the information systems and software engineering perspectives. Best websites are selected by experts and users are investigated to identify the common characteristics of them (Ivory and Hearst, 2002; Sinha et al., 2001). To empirically determine whether the content is more important than the graphics, webby award 2000 dataset is examined to differentiate the factors of the best websites from the factors of other websites (Sinha et al., 2001). Webby award evaluators use five specific criteria. The criteria include structure, content, navigation, visual design, functionality and interactivity. Although content was found to be more important than graphics, evaluation criteria cannot be considered independently (Sinha et al., 2001).

## 2.2 Website evaluation tool

In this literature, the survey summarises the usability evaluation method and proposes a new methodology (Ivory and Hearst, 2001). This new methodology, called WebTango, is introduced in previous research (Ivory, 2000). The WebTango is a quality checker tool, which proposes to help non-professional designers to develop their sites using quantitative measures of the navigational, informational and graphical aspects of a website. The usability evaluation approach is used in the field of the software engineering and adapted to the website usability evaluation (Brajnik, 2000a). The comparison of automated evaluation tools using consistency, adequate feedback, situational navigation, efficient navigation and flexibility as the characteristics of usability are explored is this research (Brajnik, 2000a). Website evaluation model based on the stages of a transaction in the e-market is another approach (Schubert and Selz, 1999). The three stages of the e-commerce are information stage, contract stage and payment stage. A website evaluation model is developed by applying the software quality model (Brajnik, 2002). The test method is proposed to determine whether an automated website evaluation tool uses the proper rules (Brajnik, 2000b, 2002). The validity of a set of website evaluation criteria is verified using the webby award 2000 dataset (Ivory and Hearst, 2002). Development and evaluation of a model called web-based quality function deployment is a model to link among total quality management, information technology (IT) and web engineering (Sudhahar et al., 2009).

The function of an automated website evaluation tool largely consists of capture, analysis and critique of website data (Ivory and Hearst, 2001). Capture activity records usage data. Analysis activity identifies potential usability problems. Critique activity proposes improvements for potential problems. Web accessibility initiative (WAI) of W3C classifies automated website evaluation tools into evaluation tool, repair tool and transformation tool. Analysis tools of automated website tools are divided into four types (Ivory and Hearst, 2001), which identify potential usability problems of a website. The first type of tools analyses server log-file data to identify potential problems in usage patterns. The second type of tools helps to check whether the hyper text markup language (HTML) code of a website follows the proper coding practice from a usability point of view. The third type of tools evaluates a website's usability by collecting data through a simulation of a hypothetical user experience. The fourth type of tools monitors consistency, availability and performance of a web server by stressing the server. This tool is most widely used in practice and some of the examples includes A-Prompt, watchfire bobby, UsableNet LIFT, W3C HTML Validator and National Institute of Standards and Technology (NIST). A-Prompt, Watchfire bobby, UsableNet LIFT, W3C HTML validator and NIST examine HTML to evaluate a website's usability. These tools check the conformance of web content accessibility guideline (WCAG) or Section 508 guidelines. In 1998, US government enforced the federal law rehabilitation act 508 that requires all e-ITs to be accessible by handicapped people. Therefore, every website is required to provide accessibility to all and this guideline becomes an evaluation criterion of automated website evaluation tools. Web criteria, an automated website evaluation tool evaluates the usability of a website by collecting primary statistical data through the simulation model. The primary evaluation criteria include accessibility, load time and content. NetRaker, another evaluation tool, develops an online survey which allows users to answer the survey while using the website. NetRaker does not check HTML code or analyse statistical data. Instead, it collects and analyses user survey data of a website.

Usability is recognised today as a major quality and success factor of websites. A wide range of usability evaluation techniques have been proposed and many of them are currently in use (Ivory and Hearst, 2001). They range from formal usability testing to informal usability tests conducted by usability specialists at usability labs or among real users. Automation of these techniques is desirable (Brajnik, 2000a; Cooper, 2008; Ivory and Hearst, 2001), because the techniques require usability specialists to conduct them or to analyse evaluation results, which is very resource consuming especially for very large and continuously growing websites. In addition, there is a lack of usability and accessibility experts due to an increased demand. A possible solution consists of capturing the knowledge and experience of these experts and expressing them in form of recommendations or guidelines to be reviewed and applied by designers and developers. Many automatic evaluation tools were developed to assist evaluators with guidelines review by automatically detecting and reporting ergonomic violation and making suggestions for repairing them. Representative examples of these tools include: A-Prompt, LIFT, Bobby (Cooper, 2008) and webSat (Scholtz et al., 1998). Some tools can be integrated with popular web design tools and methods. The most popular set of guidelines evaluated by most existing evaluation tools are the W3C WCAGs (http://www.w3c.org/TR/WCAG10) and Section 508 guidelines (http://www.section508.gov).

### *2.3 Evaluation method for decision-making*

Analytical hierarchy process (AHP) is a popular model to aggregate multiple criteria for decision-making (Yuen and Lau, 2008). Examples for the application of the AHP are: approaches in customer-driven product design process (Lin et al., 2008), bridge risk assessment (Wang et al., 2008), appropriate methodology for evaluating and ranking potential suppliers (Levary, 2008), determine optimal plant and distribution centre locations in a supply chain with special focus on the operational efficiencies of the distribution centres (Zahir and Sarker, 2010), determine the best combination of weighting-scaling methods for single and multiple decision-makers using the weighted-sum decision-making model (Velazquez et al., 2010). Examples for the adequate application of the fuzzy AHP are, amongst others, the assessment of water management plans (Srdjevic and Medeiros, 2008); safety management in production (Dagdeviren and Yüksel, 2008); personnel selection (Güngör et al., 2009) and weapon selection (Dagdeviren et al., 2009), optimum underground mining method selection (Masoud Zare et al., 2009) and shipping registry selection (Metin et al., 2009).

### *2.4 Quality standard*

Every webpage design has their own characteristics and these characteristics have drawbacks and benefits. There is a mechanism for measuring the effects of the webpage component towards the performance and quality of website. This mechanism will measure size, component and time needed by the client for downloading a website. The effective load on the system and has implications on quality of service (Mohan et al., 2009). A model and an algorithm developed to evaluate the effects of resource locking on the performance of internet applications (Mohan et al., 2010). The main factors that will influence this download time are page size (bytes), number and types of component, and number of server from the accessed web. Research conducted by IBM can be used as a standard for performance measurement of quality (Amerson et al., 2001). Table 2 describes all of the criteria and quality standard that should be fulfilled by a website to be a good quality website. Tested factors consist of: average server response time, number of component per page, webpage loading time and webpage size in byte. A standard international download time for this performance can be used as a reference to categorise the tested webpage. Automation of the testing for website quality is a new chance and a new method and should be applied for testing the quality of website.

Broken links can give a bad impact for the credibility of a website. Credibility is very important in the world wide web, because transaction between customer and seller is not on the spot and the risk of fraud is several times higher. The customers would certainly choose to buy from a website that looks professional.

**Table 2** Standard of the website performance

| *Tested factor* | *Quality standard* |
|---|---|
| Average server response time | <0.5 sec |
| Number of component per page | <20 objects |
| Webpage loading time | <30 sec |
| Webpage size in byte | <64 kbytes |

*Source*: Amerson et al. (2001).

## 3 Methodology

This research is consisted of several stages, started with problem identification followed by research procedure and data collection and ended with analysis of data. Basically, our research purpose has threefold aims:

1. to propose the new methodology for evaluating the quality of Malaysian University websites
2. to determine the best Malaysian University website based on the criteria proposed in the new methodology
3. to determine the best ranking method used to evaluate website quality.

This research examined the selected Malaysian University websites: USM, UKM, UPM, UUM and UTP. The data of quality website from Malaysian University websites was taken more than 30 trials on various occasions on the different period of time. This data has been taken from 29 March 2009 until 20 May 2009. Using website diagnostic tools and four methods proposed (linear weightage model (LWM), AHP, fuzzy analytical hierarchy process (FAHP) and new hybrid model (NHM)), the aims of this research were explored. Data was analysed using non-parametric statistical test. To analyse whether there are differences among the ranking composition methods, we used the Friedman test. When the null hypothesis is rejected by the Friedman test, we can proceed with a *post-hoc* test to detect which differences among the methods are significant using Bonferroni's/Dunn's multiple comparison technique. All of the data for this research were taken using PC with specification: processor pentium mobile 740, using local area network internet connection with average bandwidth 60 kbps.

### 3.1 Web diagnostic tools

We used a number of widely available web diagnostic tools online; thus, we used widely available website performance tool and webpage speed analyser online service (http://www.websiteoptimisation.com). List of performance measured and reported by this service include total size, number of objects (HTML, images, cascading style sheet, scripts) and download times on a 56.6 kbps connection. Another available webpage online tool that we used for testing quality was http://validator.w3.org/checklink, which was utilised to monitor broken links in the HTML code of the portals. The W3C's HTML validator website (http://validator.w3.org) was used to validate the HTML code of the portals. This standard was set up by the W3C, the main international standards organisation for the world wide web. A website tool for measuring link popularity website (www.linkpopularity.com) was used to determine the amount and quality of links that are made to a single website from many websites, based on the page-rank analysis.

This research was also conducted using accessibility online software for testing whether the webpage tested already fulfil the criteria to be accessed by people with disabilities. This software has an ability to conduct an online test for webpage referring to the criteria set-up by W3C-WCAG. WCAG is part of a series of web accessibility guidelines published by the W3C's WAI. During this research, we used Tawdis software tester that can cover almost 90% of the items demanded by WCAG.

### 3.2 Sample data

To get the data for this research, we examined Malaysian University websites from five universities. The Malaysian University websites were not randomly selected, but a careful process was undertaken. Rather than selecting any generic Malaysian University websites, this research attempted to evaluate the leading universities that are considered to be leaders in the area of IT implementation based on evaluation results by webometrics. By doing such an approach, it was felt that measures of 'best practices' could emerge. The five Malaysian University websites and their web address were: USM (http://www.usm.my), UPM (http://www.upm.edu.my), UKM (http://www.ukm.my), UUM (http://www.uum.edu.my) and UTP (http://www.utp.edu.my).

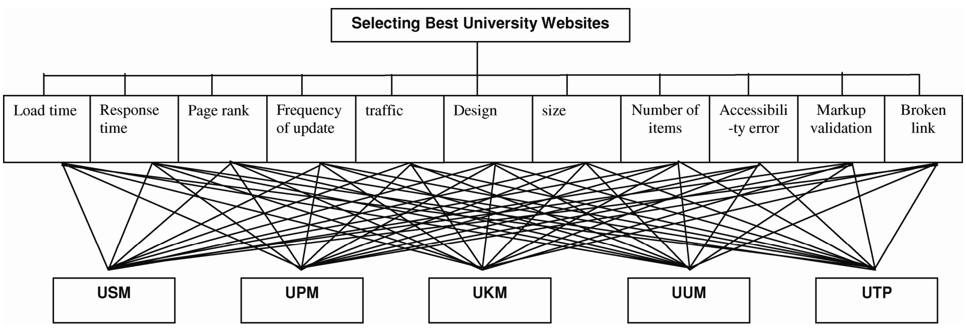
### 3.3 Linear weightage model

This model is very easy and mostly depending upon decision-makers' judgement as they have to assign weights to the criteria that involve in decision-making process. In most cases, there are some criteria considered as more important than others, such as load time, response time, traffic, page rank and broken link. Decision-makers should assign weight to each individual criterion to determine the relative importance of each one. The weight plays a vital role in decision-making process and extremely affects the final decision. First of all, decision-makers have to identify all criteria that involved in a certain process before performing any other steps. After identifying all the criteria related to website selection decision, decision-makers have to determine threshold for each criterion. In fact, a threshold can be divided into two types, i.e. maximum and minimum. To establish a threshold to criterion, decision-makers should classify all criteria into two groups. The first group known as 'larger is better' while the other known as 'smaller is better'. The load time, response time, markup validation number error and broken link can be categorised as 'smaller is better' and the threshold for this type of criteria must be maximum. On the other hand, other criteria can be considered as 'larger is better' such as traffic, page rank, frequency of update and design optimisation where the threshold must be minimum.

### *3.4 Analytic hierarchy process*

AHP was originally designed by Saaty (1980) to solve complicated multi-criteria decision problem, e.g. was used to help engineers determine the manufacturing process yield quickly and effectively (Chang et al., 2007), beside that AHP is also appropriate whenever a target is obviously declared and a set of relevant criteria and alternatives are offered (Ozden and Karpak, 2005). AHP has been proposed for determining the best website to support researcher through the decision-making activity, which aims to determine the best website among pool of university websites. In AHP, the problems are usually presented in a hierarchical structure and the decision-maker is guided throughout a subsequent series of pair-wise comparisons to express the relative strength of the elements in the hierarchy. In general, the hierarchy structure encompasses of three levels, where the top level represents the goal and the lowest level has the website under consideration. The intermediate level contains the criteria under which each website is evaluated. Evaluation cannot survive without comprehensive quality factor identification and evaluation. The AHP methodology can be demonstrated by applying it to the quality factors on the university website problem. Construction of the hierarchy is the first step in the problem-solving process. In this case (Figure 1), the goal of an AHP and also FAHP decision is to select the best university website during the first level. Load time, response time, page rank, frequency of update, traffic, design, size, number of items, accessibility error, markup validation and broken link are the evaluation criteria during the second level of the hierarchy.

**Figure 1** AHP/FAHP model of Malaysian University websites




### *3.5 Fuzzy analytic hierarchy process*

In 1965, Lotfi A. Zadeh introduced a new approach called fuzzy logic. Fuzzy sets and fuzzy logic are powerful mathematical tools for modelling: nature and humanity, uncertain systems in industry and facilitators for common-sense reasoning in decision-making in the absence of complete and precise information. One of the implementation examples of FAHP in operation research is used in selection of enterprise resource planning systems (Cebeci, 2009). The role is significant when applied to complex phenomena not easily described by traditional mathematical methods, especially when the goal is to find a good approximate solution (BojadZier and Bojadzier, 1995). The values of fuzzy logic are ranging from 0 to 1 for showing the membership of the objects in a fuzzy set. Complete non-membership is represented by 0 and complete membership is represented by 1. Values between 0 and 1 represent intermediate degrees of membership. Weight parameter for AHP and FAHP is depicted in Table 3.

Decimal judgements, like 3.5, are allowed for fine-tuning and judgements greater than 9 may be entered though it is suggested to be avoided.

### *3.6 Hybrid method*

Hybrid method combines two previous evaluation methods used before. This model is a combination between LWM and FAHP and assigns weights to the criteria using FAHP process.

### *3.7 Reliability and validity*

After the data was collected, they were then organised and analysed. The data was analysed by using non-parametric statistical test. To analyse whether there are differences among the ranking composition methods, we used the Friedman test (Demšar, 2006). When the null hypothesis is rejected by the Friedman test, we can proceed with a *post-hoc* test to detect which differences among the methods are significant. To answer this problem, we used Bonferroni's/Dunn's multiple comparison technique (Neave and Worthington, 1989). The Bonferroni *t* statistic is used to investigate dependent comparisons among means. This test is only good for investigating the difference between two means (i.e. cannot compare groups LWM and AHP vs. groups FAHP and hybrid). The Bonferroni *t* test is the same as a normal pair-wise comparison (*t* test), but the critical value is different.

**Table 3** Each of membership functions' parameter for AHP/FAHP

| | *Fuzzy AHP* | | | *AHP* |
|---|---|---|---|---|
| *Linguistic expressions* | $a_l$ | $a_l$ | $a_l$ | *A* |
| Equal | 1 | 1 | 2 | 1 |
| Equal – moderate | 1 | 2 | 3 | 2 |
| Moderate | 2 | 3 | 4 | 3 |
| Moderate – fairly strong | 3 | 4 | 5 | 4 |
| Fairly strong | 4 | 5 | 6 | 5 |
| Fairly strong – very strong | 5 | 6 | 7 | 6 |
| Very strong | 6 | 7 | 8 | 7 |
| Very strong – absolute | 7 | 8 | 9 | 8 |
| Absolute | 8 | 9 | 9 | 9 |

## 4 Result and discussion

First column in Table 4 shows the criteria of the quality website. The criteria involve in the website selection process using proposed model are load time (A), response time (B), page rank (C), frequency of update (D), traffic (E), design optimisation (F), size (G), number of items (H), accessibility error (I), markup validation (J) and broken link (K). The second column shows the measurement unit and the rest of the columns represent the Malaysian Universities performance value.

**Table 4** Original data

| *Criteria* | *Measurement unit* | *USM* | *UPM* | *UKM* | *UUM* | *UTP* |
|---|---|---|---|---|---|---|
| A | Second | 95.51 | 85.23 | 3.59 | 12.04 | 97.58 |
| B | Second | 2.40 | 2.05 | 2.33 | 0.73 | 1.85 |
| C | Number | 778.00 | 844.00 | 377.00 | 313.00 | 152.00 |
| D | Number | 60.00 | 60.00 | 30.00 | 60.00 | 30.00 |
| E | Number | 185,700.00 | 377,300.00 | 359,000.00 | 174,600.00 | 90,400.00 |
| F | Percentage | 29.50 | 39.00 | 30.00 | 26.50 | 63.50 |
| G | Number | 456,135.00 | 381,465.00 | 16,025.00 | 41,366.00 | 478,578.00 |
| H | Number | 23.00 | 46.00 | 2.00 | 19.00 | 11.00 |
| I | Number | 26.00 | 42.00 | 9.00 | 0.00 | 5.00 |
| J | Number | 158.00 | 234.00 | 20.00 | 2.00 | 86.00 |
| K | Number | 1.00 | 19.00 | 3.00 | 0.00 | 1.00 |

Results of the websites quality test based on load time, response time, page rank, frequency of update, traffic, design optimisation, size, number of items, accessibility error, markup validation and broken link are also displayed in Table 4. The data in Table 4 shows that most of the Malaysian University websites cannot fulfil the criteria as a high quality performance website as referred in Table 2. Most of the server response, load times, size and number of items exceeded the value standardised by IBM, except UKM website in load time, size and number of items criteria. Implementation of the W3C's HTML validator highlighted that none of Malaysian University websites had HTML 4.01 valid entry page, most of them did not have DOCTYPE declarations. Consequences of this problem will be on the portability and development of the website. In term of broken link, four Malaysian University websites or 80% of the samples had a broken link.

After determining the attributes and performance results, the next step in the evaluation process was to perform a comparison of each attribute. The preference criteria matrix was obtained to compare each criterion against the others. There are four models used in this research: LWM, AHP, FAHP and NHM (combination between LWM and FAHP). Table 5 presents the weights of Malaysian University website associated with each of the website quality criteria using LWM model. The load time, response time, markup validation number error and broken link can be categorised as ‘smaller is better’ and the threshold for this type of criteria must be maximum. On the other hand, other criteria can be considered as ‘larger is better’ such as traffic, page rank, frequency of update and design optimisation where threshold must be minimum. Once the attribute is considered as maximum type of threshold, formula 1 should be used.

$$ws_{min} = \frac{max - website}{max - min} \quad (1)$$

$$ws_{min} = \frac{website - min}{max - min} \quad (2)$$

where

$ws_{max}$ = specific website value that has maximum type of threshold with respect to a particular attribute/criterion.

$ws_{max}$ = specific website value that has minimum type of threshold with respect to a particular attribute/criterion.

Specific website = specific website that is considered at the time.

max = maximum value of particular attribute/criteria among all websites.

min = minimum value of the same attribute among the whole websites.

**Table 5** Final result for Malaysian University website (LWM)

| *Website/criteria* | *USM* | *UPM* | *UKM* | *UUM* | *UTP* | *Weight* |
|---|---|---|---|---|---|---|
| A (load time) | 0.02 | 0.13 | 1.00 | 0.91 | 0.00 | 0.16 |
| B (response time) | 0.00 | 0.21 | 0.04 | 1.00 | 0.33 | 0.14 |
| C (page rank) | 0.90 | 1.00 | 0.33 | 0.23 | 0.00 | 0.12 |
| D (frequency of update) | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.11 |
| E (traffic) | 0.33 | 1.00 | 0.94 | 0.29 | 0.00 | 0.11 |
| F (design optimisation) | 0.08 | 0.34 | 0.09 | 0.00 | 1.00 | 0.11 |
| G (size) | 0.05 | 0.21 | 1.00 | 0.95 | 0.00 | 0.09 |
| H (number of items) | 0.52 | 0.00 | 1.00 | 0.61 | 0.80 | 0.07 |
| I (accessibility error) | 0.38 | 0.00 | 0.79 | 1.00 | 0.88 | 0.05 |
| J (markup validation) | 0.33 | 0.00 | 0.92 | 1.00 | 0.64 | 0.04 |
| K (broken link) | 0.95 | 0.00 | 0.84 | 1.00 | 0.95 | 0.02 |
| Sum | 0.35 | 0.44 | 0.56 | 0.68 | 0.29 | |
| Rank | 4 | 3 | 2 | 1 | 5 | |

The idea of using formulas 1 and 2 is extremely valuable because they provide a method that enables the comparisons among decision criteria. Usually decision criteria have different units of measure so any comparisons among those criteria are not logically acceptable. By using the data normalisation concept, which is represented in formulas 1 and 2, all the criteria will be having weights instead of variety of measurement units and then the comparisons can simply be made. When all values of the criteria matrix are calculated, series of calculations should be achieved by multiplying weights $W_i$ of criteria by the whole values $X_i$ within the matrix. The total score should also be calculated using formula 3 for each specific website which represents the specific websites' scores. The final decision table includes a total score for each website and the one who gains the highest score is recommended as the best website overall.

$$\text{Total score} = \frac{\sum W_i X_i}{\sum W_i} \tag{3}$$

We gave every criterion with the appropriate weight depending on the significant to the total quality of website: load time (9), response time (8), page rank (7), frequency of update (6), traffic (6), design optimisation (6), size (5), number of items (4), accessibility error (3), markup validation (2) and broken link (1).

After conducting some calculations during this evaluation process, the last step in this procedure was computing the final score of each website. Then, getting the sum of each column and the sum represents the score of each single website. Table 5 depicts the final scores of websites based on LWM evaluation method. The most important thing with regards to the final results, the website which has the highest score is suggested as the best website for LWM model. In accordance with the results generated by the proposed model, UUM website has the highest score of 0.68 in comparison with the rest of other websites. As a result, the proposed LWM model rank for university website is: UUM (score: 0.68), UKM (score: 0.67), UPM (score: 0.44), USM (score: 0.35) and the last rank is UTP (score: 0.29). Table 6 presents the weights of university website associated with each of the website quality criteria based on AHP model.

**Table 6** Weight of criteria and website AHP

| *Website/criteria* | *USM* | *UPM* | *UKM* | *UUM* | *UTP* | *Weight* |
|---|---|---|---|---|---|---|
| A (load time) | 0.053 | 0.095 | 0.508 | 0.307 | 0.037 | 0.270 |
| B (response time) | 0.042 | 0.128 | 0.064 | 0.553 | 0.212 | 0.197 |
| C (page rank) | 0.275 | 0.475 | 0.121 | 0.092 | 0.038 | 0.148 |
| D (frequency of update) | 0.286 | 0.286 | 0.095 | 0.286 | 0.047 | 0.107 |
| E (traffic) | 0.116 | 0.464 | 0.303 | 0.080 | 0.037 | 0.076 |
| F (design optimisation) | 0.076 | 0.254 | 0.119 | 0.049 | 0.502 | 0.052 |
| G (size) | 0.053 | 0.102 | 0.505 | 0.305 | 0.036 | 0.042 |
| H (number of items) | 0.077 | 0.033 | 0.489 | 0.141 | 0.260 | 0.042 |
| I (accessibility error) | 0.080 | 0.032 | 0.193 | 0.443 | 0.252 | 0.030 |
| J (markup validation) | 0.071 | 0.032 | 0.260 | 0.480 | 0.156 | 0.021 |
| K (broken link) | 0.237 | 0.029 | 0.124 | 0.373 | 0.237 | 0.016 |

The final score obtained for each website across each criterion was calculated by multiplying the weight of each criterion with the weight of each website. Website which has the highest score is suggested as the best website and decision-maker may consider that one as the best decision choice. Generally, AHP has the following four steps:

1 Define an unstructured problem and determine its goal.

2 Structure the hierarchy from the top (objectives from a decision-maker's viewpoint) through intermediate levels (criteria on which subsequent levels depend) to the lowest level, which typically contains a list of alternatives.

3 Employ a pair-wise comparison approach. Fundamental scale for pair-wise comparisons developed to solve this problem (Saaty, 1980). The pair-wise comparison matrix A, in which the element $a_{ij}$ of the matrix is the relative importance of the *i*th factor with respect to the *j*th factor, could be calculated as:

$$A = \left[ a_{ij} \right] = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix} \quad (4)$$

4 There are $n(n-1)/$ judgements required for developing the set of matrices in step 3. Reciprocals are automatically assigned to each pair-wise comparison, where *n* is the matrix size.

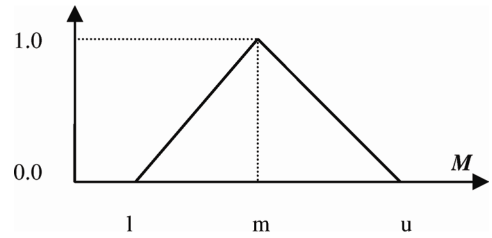
Table 7 depicts the final scores of websites based on AHP evaluation method.

**Table 7** Final result evaluation (AHP)

| *Website/criteria* | *USM* | *UPM* | *UKM* | *UUM* | *UTP* |
|---|---|---|---|---|---|
| A (load time) | 0.014 | 0.026 | 0.137 | 0.083 | 0.010 |
| B (response time) | 0.008 | 0.025 | 0.013 | 0.109 | 0.042 |
| C (page rank) | 0.041 | 0.070 | 0.018 | 0.014 | 0.006 |
| D (frequency of update) | 0.031 | 0.031 | 0.010 | 0.031 | 0.005 |
| E (traffic) | 0.009 | 0.035 | 0.023 | 0.006 | 0.003 |
| F (design optimisation) | 0.004 | 0.013 | 0.006 | 0.003 | 0.026 |
| G (size) | 0.002 | 0.004 | 0.021 | 0.013 | 0.002 |
| H (number of items) | 0.003 | 0.001 | 0.021 | 0.006 | 0.011 |
| I (accessibility error) | 0.002 | 0.001 | 0.006 | 0.013 | 0.008 |
| J (markup validation) | 0.001 | 0.001 | 0.005 | 0.010 | 0.003 |
| K (broken link) | 0.004 | 0.000 | 0.002 | 0.006 | 0.004 |
| Sum | 0.120 | 0.208 | 0.262 | 0.293 | 0.118 |
| Rank | 4 | 3 | 2 | 1 | 5 |

In accordance with the results generated by the proposed model, UUM website has the highest score of 0.293 in comparison with the rest of other websites. As a result, the proposed AHP model rank for university website is: UUM (score: 0.293), UKM (score: 0.262), UPM (score: 0.208), USM (score: 0.120) and the last rank is UTP (score: 0.118). Table 8 presents the weights of Malaysian University website associated with each of the website quality criteria based on FAHP model. Fuzzy numbers are the special classes of fuzzy quantities. A fuzzy number is a fuzzy quantity *M* that represents a generalisation of a real number *r*. Intuitively, $M(x)$ should be a measure of how better $M(x)$ 'approximates' *r*. A fuzzy number *M* is a convex normalised fuzzy set. A fuzzy number is characterised by a given interval of real numbers, each with a grade of membership between 0 and 1 (Deng, 1999). A triangular fuzzy number (TFN), *M* is shown in Figure 2.

**Figure 2** A triangular fuzzy number, $\tilde{M}$




TFNs are described by three real numbers, expressed as (*l*, *m*, *u*). The parameters *l*, *m* and *u* indicate the smallest possible value, the most promising value and the largest possible value, respectively, that describe a fuzzy event. Their membership functions are described as:

$$\mu\left(\frac{x}{\tilde{M}}\right) = \begin{cases} 0, & x < l \\ \dfrac{(x-l)}{(m-l)}, & l \le x \le m \\ \dfrac{(u-x)}{(u-m)}, & m \le x \le u \\ 0, & x > u \end{cases} \tag{5}$$

In applications, it is easy to work with TFNs because of their simple computation and they are useful in promoting representation and information processing in a fuzzy environment. In this research, implementation of TFNs in the FAHP is adopted. We have to deal with fuzzy numbers when we want to use fuzzy sets in applications. In this section, three important operations used in this research are illustrated (Tang and Beynon, 2005). If we define, two TFNs *A* and *B* by the triplets $A = (l_1, m_1, u_1)$ and $B = (l_2, m_2, u_2)$. In this research, the extent FAHP is used. Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$ an object set and $G = \{g_1, g_2, g_3, \ldots, g_n\}$ be a goal set. According to the method of Chang's extent analysis, each object is taken and extent analysis for each goal performed, respectively. Therefore, *m* extent analysis values for each object can be obtained with the following signs:

$$M_{gi}^1, M_{gi}^2, \ldots, M_{gi}^m, \quad i = 1, 2, \ldots, n$$

where $M_{gi}^j (j = 1, 2, \ldots, m)$ all are TFNs. The steps of Chang's extent analysis can be given as in the following:

*Step 1* The value of fuzzy synthetic extent with respect to the *i*th object is defined as:

$$S_i = \sum_{j=1}^{m} M_{gi}^{j} \otimes \left[ \sum_{i=1}^{n} \sum_{j=1}^{m} M_{gi}^{j} \right]^{-1} \tag{6}$$

*Step 2* As $\tilde{M}_1 = (l_1, m_1, u_1)$ and $\tilde{M}_2 = (l_2, m_2, u_2)$ are two TFNs, the degree of possibility of $M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$ is defined as:

$$V\left(\tilde{M}_2 \geq \tilde{M}_1\right) = \sup_{y \geq x} \left[ \min\left(\mu_{\tilde{M}_1}(x), \mu_{\tilde{M}_2}(y)\right) \right] \tag{7}$$

and can be equivalently expressed as follows:

$$V\left(\tilde{M}_2 \geq \tilde{M}_1\right) = hgt\left(\tilde{M}_1 \cap \tilde{M}_2\right) = \mu_{M_2}(d) \tag{8}$$

$$= \begin{cases} 1, & \text{if } m_2 \geq m_1 \\ 0, & \text{if } l_1 \geq u_2 \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & \text{otherwise} \end{cases} \tag{9}$$

*Step 3* The degree possibility for a convex fuzzy number to be greater than *k* convex fuzzy $M_i$ $(i = 1, 2, k)$ numbers can be defined by:

$$\begin{aligned} V\left(M > M_1, M_2, \ldots, M_k\right) &= V\left[\left(M > M_1\right) \text{and} \left(M > M_2\right) \text{and} \left(M > M_k\right)\right] \\ &= \min V\left(M \geq M_i\right), i = 1, 2, 3, \ldots, k \end{aligned} \tag{10}$$

Assume that $d(A_i) = \min V\left(S_i \geq S_k\right)$ for $k = 1, 2, \ldots, n;\ k \neq i$.

Then, the weight vector is given by:

$$W' = \left(d'\left(A_1\right), d'\left(A_2\right), \ldots, d'\left(A_n\right)\right)^{\mathrm{T}} \tag{11}$$

where $A_i = (i = 1, 2, \ldots, n)$ are *n* elements.

**Figure 3** The intersection between $M_1$ and $M_2$

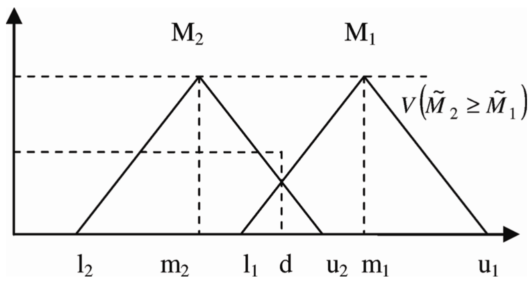


Figure 3 illustrates Equation (9) where *d* is the ordinate of the highest intersection point *D* between $\mu_{M_1}$ and $\mu_{M_2}$ to compare $M_1$ and $M_2$, we need both the values of $V(M_1 \geq M_2)$ and $V(M_2 \geq M_1)$.

*Step 4* Via normalisation, the normalised weight vectors are:

$$W = \left(d\left(A_1\right), d\left(A_2\right), \ldots, d\left(A_n\right)\right)^{\mathrm{T}} \tag{12}$$

where *W* is a non-fuzzy number.

Table 8 presents the weights of Malaysian University websites associated with each of the website quality criteria based on FAHP model.

Table 9 depicts the final scores of websites. The most important thing with regards to the final results, the website which has the highest score is suggested as the best website for the proposed FAHP model.

**Table 8** Weight criteria and website (FAHP)

| *Website/criteria* | *USM* | *UPM* | *UKM* | *UUM* | *UTP* | *Weight* |
|---|---|---|---|---|---|---|
| A (load time) | 0.000 | 0.000 | 0.613 | 0.387 | 0.000 | 0.377 |
| B (response time) | 0.000 | 0.000 | 0.000 | 0.933 | 0.067 | 0.291 |
| C (page rank) | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.216 |
| D (frequency of update) | 0.333 | 0.333 | 0.333 | 0.333 | 0.333 | 0.114 |
| E (traffic) | 0.000 | 0.549 | 0.451 | 0.000 | 0.000 | 0.003 |
| F (design optimisation) | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 |
| G (size) | 0.000 | 0.000 | 0.843 | 0.157 | 0.000 | 0.000 |
| H (number of items) | 0.000 | 0.000 | 0.683 | 0.000 | 0.317 | 0.000 |
| I (accessibility error) | 0.000 | 0.000 | 0.176 | 0.488 | 0.337 | 0.000 |
| J (markup validation) | 0.000 | 0.000 | 0.314 | 0.686 | 0.000 | 0.000 |
| K (broken link) | 0.257 | 0.000 | 0.000 | 0.524 | 0.219 | 0.000 |

**Table 9** Final result (FAHP)

| *Website/criteria* | *USM* | *UPM* | *UKM* | *UUM* | *UTP* |
|---|---|---|---|---|---|
| A (load time) | 0.000 | 0.000 | 0.231 | 0.146 | 0.000 |
| B (response time) | 0.000 | 0.000 | 0.000 | 0.272 | 0.019 |
| C (page rank) | 0.000 | 0.216 | 0.000 | 0.000 | 0.000 |
| D (frequency of update) | 0.038 | 0.038 | 0.038 | 0.038 | 0.038 |
| E (traffic) | 0.000 | 0.002 | 0.001 | 0.000 | 0.000 |
| F (design optimisation) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| G (size) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| H (number of items) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| I (accessibility error) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| J (markup validation) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| K (broken link) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Total | 0.038 | 0.256 | 0.270 | 0.455 | 0.057 |
| Rank | 5 | 3 | 2 | 1 | 4 |

The website which has the highest score is suggested as the best website for the proposed FAHP model. UUM website has the highest score of 0.455 in comparison with the rest of university websites. As a result, the proposed FAHP model rank for university website is: UUM (score: 0.455), UKM (score: 0.270), UPM (score: 0.256), UTP (score: 0.057) and the last rank is USM (score: 0.038). Table 10 presents the weights of university websites associated with each of the website quality criteria based on hybrid model.

Table 10 is derived from Tables 5 and 8. Hybrid method combines two previous evaluation methods used before. This model has to assign weights to the criteria that involve in decision-making process. Weight for alternative is taken from FAHP process and weight for criteria is taken from LWM. Applying hybrid model between FAHP and LWM approach for website evaluation has resulted in significant reduction of computation, raised the overall speed and effectiveness with respect to the underlying methodology and ultimately enabled more efficient and significantly procedure compared with other methods.

Table 11 depicts the final scores of websites. The most important with regards to the final results, the website which has the highest score is suggested as the best website for the proposed hybrid model. In accordance with the results generated by the proposed model, UUM website has the highest score of 0.799 in comparison with the rest of university websites. As a result, the proposed hybrid model rank for university website is: UUM (score: 0.799), UKM (score: 0.461), UPM (0.441), UPM (score: 0.318) and the last rank is UTP (score: 0.095).

Table 12 depicts the final scores of Malaysian University websites based on four evaluation methods, UUM has the highest in score in LWM, AHP, FAHP and NHM compared with the rest of university websites. Inconsistency occurred for the FAHP model, different with other three models, because for USM rank 5 and UTP rank 4. To analyse whether there is differences among the ranking composition methods, we used the Friedman test (Demšar, 2006). When the null hypothesis is rejected by the Friedman test, we can proceed with a *post-hoc* test to detect which differences among the methods are significant and this procedure displayed in Table 13.

**Table 10** Maxium–minimum criteria (hybrid)

| *Website/criteria* | | *USM* | *UPM* | *UKM* | *UUM* | *UTP* | *Weight* |
|---|---|---|---|---|---|---|---|
| A (load time) | Max | 0.02 | 0.13 | 1.00 | 0.91 | 0.00 | 0.377 |
| B (response time) | Max | 0.00 | 0.21 | 0.04 | 1.00 | 0.33 | 0.291 |
| C (page rank) | Min | 0.90 | 1.00 | 0.33 | 0.23 | 0.00 | 0.216 |
| D (frequency of update) | Min | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.114 |
| E (traffic) | Min | 0.33 | 1.00 | 0.94 | 0.29 | 0.00 | 0.003 |
| F (design optimisation) | Min | 0.08 | 0.34 | 0.09 | 0.00 | 1.00 | 0.000 |
| G (size) | Max | 0.05 | 0.21 | 1.00 | 0.95 | 0.00 | 0.000 |
| H (number of items) | Max | 0.52 | 0.00 | 1.00 | 0.61 | 0.80 | 0.000 |
| I (accessibility error) | Max | 0.38 | 0.00 | 0.79 | 1.00 | 0.88 | 0.000 |
| J (markup validation) | Max | 0.33 | 0.00 | 0.92 | 1.00 | 0.64 | 0.000 |
| K (broken link) | Max | 0.95 | 0.00 | 0.84 | 1.00 | 0.95 | 0.000 |

**Table 11** Final result for Malaysian University website hybrid model

| *Criteria* | | *USM* | *UPM* | *UKM* | *UUM* | *UTP* |
|---|---|---|---|---|---|---|
| A | Max | 0.008 | 0.050 | 0.377 | 0.343 | 0.000 |
| B | Max | 0.000 | 0.060 | 0.011 | 0.291 | 0.095 |
| C | Min | 0.195 | 0.216 | 0.070 | 0.050 | 0.000 |
| D | Min | 0.114 | 0.114 | 0.000 | 0.114 | 0.000 |
| E | Min | 0.001 | 0.003 | 0.003 | 0.001 | 0.000 |
| F | Min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| G | Max | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| H | Max | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| I | Max | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| J | Max | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| K | Max | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Total | | 0.318 | 0.441 | 0.461 | 0.799 | 0.095 |
| Rank | | 4 | 3 | 2 | 1 | 5 |

**Table 12** Final result for e-Malaysian University websites performance

| *Method* | *USM* | *UPM* | *UKM* | *UUM* | *UTP* |
|---|---|---|---|---|---|
| LWM | 0.35(4) | 0.44(3) | 0.56(2) | 0.68(1) | 0.29(5) |
| AHP | 0.120(4) | 0.208(3) | 0.262(2) | 0.293(1) | 0.118(5) |
| FAHP | 0.038(5) | 0.256(3) | 0.270(2) | 0.455(1) | 0.057(4) |
| Hybrid | 0.318(4) | 0.441(3) | 0.461(2) | 0.799(1) | 0.095(5) |

**Table 13** Malaysian University websites ranking based on method

| *Original data* | | | | | | *Ranked data* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *LWM* | *AHP* | *FAHP* | *Hybrid* | | | *LWM* | *AHP* | *FAHP* | *Hybrid* |
| 1 | USM | 0.352 | 0.120 | 0.038 | 0.318 | 1 | USM | 4 | 2 | 1 | 3 |
| 2 | UPM | 0.437 | 0.208 | 0.256 | 0.441 | 2 | UPM | 3 | 1 | 2 | 4 |
| 3 | UKM | 0.558 | 0.262 | 0.270 | 0.461 | 3 | UKM | 4 | 1 | 2 | 3 |
| 4 | UUM | 0.680 | 0.293 | 0.455 | 0.799 | 4 | UUM | 3 | 1 | 2 | 4 |
| 5 | UTP | 0.293 | 0.118 | 0.057 | 0.095 | 5 | UTP | 4 | 3 | 1 | 2 |
| | | | | | | | Sum of rank ($SR_i$) | 18 | 8 | 8 | 16 |

To check the ranking, note that the sum of the four rank sums is 18 + 8 + 8 + 16 = 50 and that the sum of the *c* numbers in a row is $(c(c+1))/2$. However, there are *r* rows, so we must multiply the expression by *r*. So, we have $\sum SR_i = (rc(c+1))/2 = (5(4)(5))/2 = 50$.

Now, compute the Friedman statistic:

$$\chi_F^2 = \left[ \frac{12}{rc(c+1)} \sum_i \left( SR_i^2 \right) \right] - 3r(c+1)$$

$$\chi^2 = \left[ \frac{12}{(5)(4)(5)} \left( (18)^2 + (8)^2 + (8)^2 + (16)^2 \right) \right] - 3(5)(5) = \left[ \frac{12}{100}(708) \right] - 75 = 9.96$$

If we find the place on the Friedman Table for 4 columns and 5 rows, we find that the *p*-value for $\chi_F^2 = 9.96$ is 0.0185. Since the *p*-value is below $\alpha = 0.05$, reject the null hypothesis. Since the computed Friedman statistic is greater than 7.815, the upper-tail critical value under the chi-square distribution having $c - 1 = 3$ degrees of freedom (table Friedman), the null hypothesis is rejected at the 0.05 level of significant. We conclude that there are significant differences (as perceived by the raters) with respect to the rating produced at the four evaluation model. Naturally, we must now determine which methods are different from one another. To answer this question, we use Bonferroni *t*/Dunn's multiple comparison technique (Neave and Worthington, 1989). Using this method, we test $p = 12k(k - 1)$ hypotheses of the form:

$H(i, j)_0$: There is no difference in the mean average correlation coefficients between methods *i* and *j*.

$H(i, j)_1$: There is some difference in the mean average correlation coefficients between methods *i* and *j*.

Because we are allowed to make many comparisons, we have to control for family-wise error by reducing the per comparison *α* level. The overall level will be set to 0.05 and the individual per comparison *α* levels will be equal to 0.05 divided by the total number of possible comparisons. We can make a total of 4*C*2 = 6 different pair-wise comparisons between the four means. In practice, we cannot to do all of these comparisons, but remember that we will always have to set the error rate according to the total number of possible comparisons (Table 14).

**Table 14** Significance of difference between two means methods

| *LWM vs. AHP* | *LWM vs. FAHP* | *LWM vs. hybrid* | *AHP vs. FAHP* | *AHP vs. hybrid* | *FAHP vs. hybrid* |
|---|---|---|---|---|---|
| $\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{2(MS_{error})}{n}}}$ | $\frac{\bar{x}_1 - \bar{x}_3}{\sqrt{\frac{2(MS_{error})}{n}}}$ | $\frac{\bar{x}_1 - \bar{x}_4}{\sqrt{\frac{2(MS_{error})}{n}}}$ | $\frac{\bar{x}_2 - \bar{x}_3}{\sqrt{\frac{2(MS_{error})}{n}}}$ | $\frac{\bar{x}_2 - \bar{x}_4}{\sqrt{\frac{2(MS_{error})}{n}}}$ | $\frac{\bar{x}_3 - \bar{x}_4}{\sqrt{\frac{2(MS_{error})}{n}}}$ |
| $\frac{3.6 - 1.6}{\sqrt{\frac{2(0.525)}{5}}}$ | $\frac{3.6 - 1.6}{\sqrt{\frac{2(0.525)}{5}}}$ | $\frac{3.6 - 3.2}{\sqrt{\frac{2(0.525)}{5}}}$ | $\frac{1.6 - 1.6}{\sqrt{\frac{2(0.525)}{5}}}$ | $\frac{1.6 - 3.2}{\sqrt{\frac{2(0.525)}{5}}}$ | $\frac{1.6 - 3.2}{\sqrt{\frac{2(0.525)}{5}}}$ |
| =4.367 | =4.367 | =0.873 | =0 | =–3.493 | =–3.493 |

*Step 1 Calculate the t' statistics*:

General formula:

$$t^1 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\text{MS}_{\text{error}}}{4} + \frac{\text{MS}_{\text{error}}}{n}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{2(\text{MS}_{\text{error}})}{n}}}$$

$$\text{SS}_\text{T} = \sum_{i=1}^{4}\sum_{j=1}^{5} y_{ij}^2 - \frac{y^2}{N} = 150 - 125 = 25$$

$$\text{SS}_{\text{Treatment}} = \sum_{i=1}^{4} \frac{y_i^2}{n} - \frac{y^2}{N} = \frac{708}{5} - \frac{(50)^2}{20} = 141.6 - 125 = 16.6$$

$$\text{SS}_\text{E} = \text{SS}_\text{T} - \text{SS}_{\text{treatment}} = 25 - 16.6 = 8.4$$

$$\text{MS}_{\text{treatment}} = \frac{\text{SS}_{\text{treatment}}}{a-1} = \frac{16.6}{3} = 5.53$$

$$\text{MS}_\text{E} = \frac{\text{SS}_\text{e}}{[a(n-1)]} = \frac{8.4}{[4(4)]} = 0.525$$

General formula:

$$t^1 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\text{MS}_{\text{error}}}{n} + \frac{\text{MS}_{\text{error}}}{n}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{2(\text{MS}_{\text{error}})}{n}}}$$

*Step 2 Set α to the appropriate level*:

We want to keep our family-wise error at or below $\alpha = 0.05$. We have three comparisons, so our per comparison $\alpha$ will be:

$$\propto = \frac{\alpha_{\text{FW}}}{k} = \frac{0.05}{4} = 0.0125$$

$$\text{d}t = \text{d}f\text{MS}_{\text{error}} = 16$$

*Step 3 Determine significance of comparisons*:

Student's *t* tables do not contain a critical value for $\alpha = 0.0125$, so we have to look it up in the Dunn/Bonferroni *t*' table. The degrees of freedom = 16 and the number of comparison = 6. This gives a *t*' value: 3.008. Result for this test LWM vs. AHP: *t*' = 4.367 (significant), LWM vs. FAHP: *t*' = 4.367 (significant), LWM vs. hybrid: *t*' = 0.873 (insignificant), AHP vs. FAHP: *t*' = 0.00 (insignificant), AHP vs. hybrid: *t*' = – 3.493 (significant), FAHP vs. hybrid: *t*' = – 3.493 (significant). Therefore, we can conclude that NMH and LWM ranking method are significantly better than AHP and FAHP, while there is no difference between AHP compared to FAHP methods and new hybrid compared to LWM.

## 5 Conclusion

In this paper, we evaluate the quality of Malaysian University websites with the sample of five Malaysian Universities. Using a series of online diagnostic tools, we examined many dimensions of quality and each dimension was measured by specific test online. The result of this study confirmed that Malaysian University website is neglecting performance and quality criteria was standardised by IBM (Table 2). It is clear in our research that more effort is required to meet these criteria in the context of website design. This suggests that web developers responsible for Malaysian University websites should follow and encourage the use of recognised guidelines when designing website. To get results on the quality of a website, we measured sample data from Malaysian University websites and calculated load time, response time, page rank, frequency of update, traffic, design optimisation, size, number of items, accessibility error, markup validation and broken link. The proposed NHM has been implemented using LWM and FAHP to generate the weights for the criteria, which are much better and guaranteed more fairly preference of criteria, applying hybrid model between LWM and FAHP approach for website evaluation has resulted in significant acceleration of implementation, raised the overall effectiveness with respect to the underlying methodology and ultimately enabled more efficient and significantly better compared with AHP and FAHP method. Limitation of this research occurred in the number of sample size and time factor, this research used a limited sample size of 30 data and taken during a short period of observation time.

Future directions for this research are added criteria for evaluating websites quality, such as availability and security aspect, also from the cultural perspective, since culture has an impact upon a websites. Another approach also can be conducted for other service sectors like e-government website. Moreover, since the ultimate determinant of quality website is the users, future direction for this research also involves the objective and subjective views of the e-Malaysian University websites from user's perspective. This research could be repeated every year to gather time-based data (longitudinal study). For example, the progress of the initiatives and its quality level could be plotted against time to measure the progress year-to-year. Standing of the Asian University website is to be expanded with more countries because this research only analysed university in Malaysia. A wider view of sample would result in a more accurate conclusion. Best practices that could be applied from other region that is already soaring in implementation of website applications (e-government, e-commerce and university) such as European and North America need to be highlighted. This would benefit emerging economical power like Asian countries and eliminates the need to reinvent the wheel.

## References

Amerson, M., Fisher, G., Hsiung, L., Krueger, L. and Mills, N. (2001) 'Design for performance: analysis of download times for page elements suggests ways to optimize', Available at: http://www.ibm.com/developerworks/websphere/library/techarticles/hipods/perform.html.

Anagnostopoulos, K.P. and Mamanis, G. (2010) 'A portfolio optimization model with three objectives and discrete variables', *Computers & Operations Research*, Vol. 37, pp.1285–1297.

BojadZier, G. and Bojadzier, M. (1995) 'Fuzzy sets, fuzzy logic, applications', *Advanced in Fuzzy Systems – Applications and Theory*, Vol. 5. Singapore: World Scientific Publishing.

Bouch, A., Kuchnisky, A. and Bhatti, N. (2000) 'Quality is in the eye of the beholder: meeting users' requirements for internet quality of service', in T. Turner and G. Szwillus (Eds.), *Human Factors in Computing Systems.* The Hague, The Netherlands: ACM.

Brajnik, G. (2000a) 'Automatic web usability evaluation: where is the limit', *Proceedings of the Sixth Conference on Human Factors and the Web.*

Brajnik, G. (2000b) 'Automatic web usability evaluation: what needs to be done?', in Ph. Kortum and E. Kudizinger (Eds.), *Proceedings of Sixth Conference on Human Factors and the Web.*

Brajnik, G. (2002) 'Quality models based on automatic webtesting', *CHI Workshop*, Minneapolis: ACM.

Branke, J., Scheckenbach, B., Stein, M., Deb, K. and Schmeck, H. (2009) 'Portfolio optimization with an envelope-based multi-objective evolutionary algorithm', *European Journal of Operational Research*, Vol. 199, pp.684–693.

Cebeci, U. (2009) 'Fuzzy AHP-based decision support system for selecting ERP systems in textile industry by using balanced scorecard', *Expert Systems with Applications*, Vol. 36, pp.8900–8909.

Chang, C.W., Wu, C.R., Lin, C.T. and Chen, H.C. (2007) 'An application of AHP and sensitivity analysis for selecting the best slicing machine', *Computers & Industrial Engineering*, Vol. 52, pp.296–307.

Cooper, M. (2008) 'W3C web content accessibility guideline (WCAG) 2.0', Available at: http://www.w3.org/TR/2008/REC-WCAG20-20081211/.

Dagdeviren, M., Yavuz, S. and KilIn, N. (2009) 'Weapon selection using the AHP and TOPSIS methods under fuzzy environment', *Expert Systems with Applications*, Vol. 36, pp.8143–8151.

Dagdeviren, M. and Yüksel, I (2008) 'Developing a fuzzy analytic hierarchy process (AHP) model for behavior-based safety management', *Information Sciences*, Vol. 178, pp.1717–1733.

Demšar, J. (2006) 'Statistical comparisons of classifiers over multiple data sets', *The Journal of Machine Learning Research*, Vol. 7, pp.1–30.

Deng, H. (1999) 'Multicriteria analysis with fuzzy pairwise comparison', *Int. J. Approximate Reasoning*, Vol. 21, No. 3, pp.215–231.

Durkin, M. (2007) 'In search of the internet-banking customer: exploring the use of decision styles', *Int. J. Bank Marketing*, Vol. 22, pp.484–503.

Green, D. and Pearson, J.M. (2006) 'Development of a web site usability instrument based on ISO-9241-11', *The Journal of Computer Information Systems*, Vol. 47, pp.66–72.

Güngör, Z., SerhadlIoglu, G. and Kesen, S.E. (2009) 'A fuzzy AHP approach to personnel selection problem', *Applied Soft Computing*, Vol. 9, pp.641–646.

Huizingh, E. (2000) 'The content and design of web sites: an empirical study', *Information & Management*, Vol. 37, pp.123–134.

Ivory, M. and Hearst, M. (2001) 'The state of the art in automating usability evaluation of user interfaces', *ACM Computing Surveys*, Vol. 33, pp.470–516.

Ivory, M. and Hearst, M. (2002) 'Statistical profiles of highly-rated web sites', Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves. Minneapolis, MN: ACM.

Ivory, M.Y. (2000) Web TANGO: Towards Automated Comparison of Information-centric Web Site Designs. New York, NY: ACM.

Krishnamurthy, B. and Wills, C. (2006) 'Cat and mouse: content delivery tradeoffs in web access', *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland.

Levary, R.R. (2008) 'Using the analytic hierarchy process to rank foreign suppliers based on supply risks', *Computers & Industrial Engineering*, Vol. 55, pp.535–542.

Lin, M.C., Wang, C.C., Chen, M.S. and Chang, C.A. (2008) 'Using AHP and TOPSIS approaches in customer-driven product design process', *Computers in Industry*, Vol. 59, pp.17–31.

Loiacono, E. and McCoy, S. (2004) 'Web site accessibility: an online sector analysis', *Information Technology & People*, Vol. 17, pp.87–101.

Loiacono, E.T., Watson, R.T. and Goodhue, D.L. (2007) 'WebQual: an instrument for consumer evaluation of web sites', *Int. J. Electronic Commerce*, Vol. 11, pp.51–87.

Masoud Zare, N., Reza, M. and Mohammad, A. (2009) 'The application of fuzzy analytic hierarchy process (FAHP) approach to selection of optimum underground mining method for Jajarm Bauxite Mine, Iran', *Expert System Application*, Vol. 36, pp.8218–8226.

McCowen, F., Michael, N. and Bollen, J. (2005) 'The availability and persistence of web references in D-Lib magazine', *Proceedings of the 5th International Web Archiving Workshop and Digital Preservation (IWAW'05)*, Vienna.

McInerney, C. (2000) 'Establishing and maintaining trust in online systems', in M. Williams (Ed.), *The 21st National Online Meeting Information Today*. Medford, NJ: Information Today, Idea Group Publication, pp.257–270.

Metin, C., Er, I.D. and Ozok, A.F. (2009) 'Application of fuzzy extended AHP methodology on shipping registry selection: the case of Turkish maritime industry', *Expert System Application*, Vol. 36, pp.190–198.

Mohan, S., Printezis, A. and Alam, F.M. (2009) 'A framework for modelling web-based applications with resource locking', *Int. J. Operational Research*, Vol. 6, pp.289–303.

Mohan, S., Printezis, A. and Alam, M.F. (2010) 'Impact of resource locking on performance of web-based applications', *Int. J. Operational Research*, Vol. 8, pp.355–378.

Neave, H.R. and Worthington, P.L. (1989) *Distribution-Free Tests*. London: Routledge.

Olsina, L., Godoy, D., Lafuente, G.J. and Rossi, G. (2001) 'Specifying quality characteristics and attributes for websites', in S. Murugesan and Y. Deshpande (Eds.), *Web Engineering*, LNCS, 2016. Berlin: Springer, pp.266–278.

Ozden, B. and Karpak, B. (2005) 'An AHP application in vendor selection', *ISAHP 2005*, Honolulu, Hawaii, 8–10 July 2005, pp.1–21.

Page, L., Motwani, R., Brin, S. and Winograd, T. (1998) 'The anatomy of a large-scale hypertextual web search engine', *Computer Networks and ISDN Systems*, Vol. 30, Nos. 1–7, pp.107–117.

Ramsay, J., Barbesi, A. and Preece, J. (1998) 'Psychological investigation of long retrieval times on the world wide web', *Interacting with Computers*, Vol. 10, pp.77–86.

Saaty, T. (1980) The Analytic Hierarchy Process: Planning, Priority Setting, Resources Allocation. London: McGraw-Hill.

Sayar, C. and Wolfe, S. (2007) 'Internet banking market performance: Turkey versus the UK', *Int. J. Bank Marketing*, Vol. 25, pp.122–141.

Scholtz, J., Laskowski, S. and Downey, L. (1998) 'Developing usability tools and techniques for designing and testing web sites', *Proceedings of the Fourth Conference on Human Factors and the Web*, Available at: http://www.research.att.com/conf/hfweb/index.html.

Schubert, P. and Selz, D. (1999) 'Web assessment – measuring the effectiveness of electronic commerce sites going beyond traditional marketing paradigms', *Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences*, IEEE Computer Society.

Seffah, A., Donyaee, M., Kline, R. and Padda, H. (2006) 'Usability measurement and metrics: a consolidated model', *Software Quality Journal*, Vol. 14, pp.159–178.

Sinha, R., Hearst, M. and Ivory, M. (2001) 'Content or Graphics?: an empirical analysis of criteria for award-winning websites', *Proceedings of the 7th Conference on Human Factors & the Web*.

Srdjevic, B. and Medeiros, Y.D.P. (2008) 'Fuzzy AHP assessment of water management plans', *Water Resources Management*, Vol. 22, pp.877–894.

Sudhahar, C., Kumar, R.S.P., Senthil, V., Devadasan, S.R. and Murugesh, R. (2009) 'Web-based QFD: a collaborative technique for supporting the adoption of customers' voices', *Int. J. Business Information Systems*, Vol. 4, pp.360–385.

Tang, Y.C. and Beynon, M.J. (2005) 'Application and development of a fuzzy analytic hierarchy process within a capital investment study', *Journal of Economics and Management*, Vol. 1, pp.207–230.

Tiryaki, F. and Ahlatcioglu, B. (2009) 'Fuzzy portfolio selection using fuzzy analytic hierarchy process', *Information Sciences*, Vol. 179, pp.53–69.

Velazquez, M.A., Claudio, D. and Ravindran, A.R. (2010) 'Experiments in multiple criteria selection problems with multiple decision makers', *Int. J. Operational Research*, Vol. 7, pp.413–428.

Wang, Y.M., Liu, J. and Elhag, T. (2008) 'An integrated AHP-DEA methodology for bridge risk assessment', *Computers & Industrial Engineering*, Vol. 54, pp.513–525.

Yuan, J., Chi, C. and Sun, Q. (2005) 'A more precise model for web retrieval', *Proceedings of the International World Wide Web Conference*, Chiba, Japan: ACM.

Yuen, K. and Lau, H. (2008) 'Software vendor selection using fuzzy analytic hierarchy process with ISO/IEC9126', *IAENG Int. J. Computer Science*, Vol. 35, pp.267–274.

Zahir, S. and Sarker, R. (2010) 'Optimising multi-objective location decisions in a supply chain using an AHP-enhanced goal-programming model', *Int. J. Logistics Systems and Management*, Vol. 6, pp.249–266.