A comparison of Asian airlines websites quality: using a non-parametric test

P.D.D. Dominic*

Department of Computer and Information Science,
Universiti Teknologi PETRONAS,
Tronoh 31750, Perak, Malaysia
Fax: +605 365 6180
E-mail: pdddominic@yahoo.com
E-mail: dhanapal_d@petronas.com.my
*Corresponding author

Handaru Jati

Department of Electronics Engineering Education,
Yogyaharta State University,
Korangmalang 55281, Sleman, Indonesia
Fax: +62 274 554 686
E-mail: handaruj@yahoo.com

Abstract: This research conducts test to measure the quality of Asian airlines website via web diagnostic tools online. We propose a methodology for determining and evaluating the best airlines website based on many criteria of website quality, consist of linear weightage model (LWM), analytical hierarchy process, fuzzy analytical hierarchy process (FAHP) and one new hybrid model (NHM). This NHM has been implemented using combination FAHP and LWM. The result of this study confirmed most of Asian websites are neglecting performance and quality criteria. By applying hybrid model approach has resulted in significant acceleration of implementation, raised the overall effectiveness and enabled more efficient procedure. To compare ranking methods, a combination of Friedman's test and Dunn's multiple comparison procedure is adopted.

Keywords: performance; airlines; quality; web diagnostic; hybrid model.

Biographical notes: P.D.D. Dominic obtained his MSc in Operations Research in 1985, MBA from Regional Engineering College, Tiruchirappalli, India during 1991, Post Graduate Diploma in Operations Research in 2000 and completed his PhD during 2004 in the area of job shop scheduling at Alagappa University, Karaikudi, India. Currently, he is working as an Associate Professor in the Department of Computer and Information Science at Universiti 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 in Information Technology in 2005 and MBA during 2001 from Gadjah Mada University Indonesia. He completed his PhD during 2010 in the area of quality evaluation of websites at Universiti Teknologi PETRONAS in the Department of Computer and Information Science, Malaysia. Since 1999, he has held the post of Lecturer in the Department of Electronics and Informatics Engineering 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 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 and media studies) and the usability of the interface (mediated communication). Each definition of quality leads to lists of criteria about what constitutes a quality site. All of these criteria from multiple studies on web quality to form a comprehensive tool for evaluating the quality of a website that would serve to assess its trustworthiness were explained in one research (McInerney, 2000). The principle was that 'if information can pass a test of quality, it is most likely to prove trustworthy' and because of this belief, should have higher credibility. Thus the challenge is how to create a method that will guide the internet user to evaluate website very easily without any problem. The method needs a lot of time and cautious consideration. It takes more than one hour to examine a website thoroughly and apply criteria of the quality. This time dedication may be available to information professionals, but for the public user 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 scope of quality, and each measure will pertain to a particular website in varying degrees. Here are some of them: first factor is time, a credible site should be updated frequently. The information about latest update should also be included on the homepage, if the information in the website has not been updated recently, the visitor could simply know that, perhaps, the site manager has no time and really bother to update the site.

Second factor is structural, all of the parts of the website hold together and all links inside and outside the website should work well. Clear navigation, legible content, clean page layouts, simple instructions and easy search functions all of this factors contribute to a site's user friendliness. Broken links on the web page are also another factor that always downgrades the quality of website. Each page usually has references or links or connections to other pages. These may be internal or external website. Users expect each link to be valid, meaning that it leads successfully to the intended page or other resource. In 2003, the authors found that one link out of every 200 disappeared each week from the internet (McCowen et al., 2005).

The third factor is content or search engine friendliness; number of the links or link popularity is one of the page factors that search engines are looking to determine the value of the web page. Increasing in the number of internet users all over the world, online businesses are absolutely on an immense raise. To generate income through good sales, institutions need to have quality web traffic first. At the very least, website should be search engine friendly. Search engines should be able to easily extract all the content for public online and display relevant pages from website to fulfil public query through search engine. Major search engines have their own way of defining relevant search results for particular key phrase. For quality web traffic, certainly need to improve website ranking on most of the search engines. To improve website ranking, there are some steps to do, first, analyse who are target audiences and then analyse what are the keywords or phrases that target audience are using on the internet while doing web searching. Second, need to update content very regularly and this action can improve website ranking quite remarkable. This is because most of search engine algorithms give top ranking while indexing the websites if the content is updated frequently. Search engine will need a website to have at least two links pointing to their site before they will place it to their index, and the idea of this link popularity is that to increase the link popularity of a website, then this website must have large amount of high-quality content. Number of links to website improves access growth and helps to generate traffic (Page and Brin, 1998). This indicator can be used to measure the quality of website.

Fourth factor is response time and latency; a website server should respond to a browser request within certain parameters, it is found that extraneous content exists on the majority of popular pages, and that blocking this content buys a 25–30% reduction in objects downloaded and bytes, with a 33% decrease in page latency, from 2003 to 2008 the average web page grew from 93.7 K to over 312 K (Josep et al., 2007). 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). Table 1 showed the ranking of airlines based on report publish by London-based consultancy firm Skytrax in 2007 (www.airlinequality.com), the ranking are Singapore Airlines, Thai Airways, Cathay Pacific, Qatar Airways, Qantas and Malaysia Airlines.

The fifth factor is stickiness, which is the ability to ensure that the internet user sticks on the website pages for a longer period of time. A sticky website is a place that people will come to visit again. By building a condition that people like to visit again, this strategy can increase exposure to product or service hence it can create 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.

Table 1   Ranking of the airlines website based on survey

Airlines	Address	Skytrax
Singapore Airlines	www.singaporeair.com	1
Thai Airways	www.thaiair.com	2
Cathay Pacific	www.cathaypacific.com	3
Qatar Airways	www.qatarairways.com	4
Qantas	www.qantas.com.au	5
Malaysia Airlines	www.malaysiaairlines.com	6

The sixth factor is design, not only does a site needs to make sense visually, but also it should read the same for all web browsers (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 an important impact in service industries and fundamentally shapes how services are delivered (Durkin, 2007). There are so many factors influence the performance of the web and most of it is outside the control of website designer. Download time of the website will be determined by web page design, web server, hardware of the client, software configuration and characteristics of the internet router which connect users and the websites. One of the research finding mention that website which has slow download time less attractive compare than website with faster download time (Ramsay et al., 1998). In the recent time, the average time of the connection speed is 5 kBps (kilobytes per second). This facts give an implication that one web page with 40 kB page size will be downloaded during 8 sec. This matter in accordance with the 'eight second rule', this 8 sec is a normal time for loading web page and will be tolerable by the user. This results are supported by many research, results mentioned that mean of tolerable download time in the user side is 8.57 with standard deviation 5.9 sec (Bouch et al., 2000). Providing information related with waiting time is very important for user. For the long download time, it is better to provide information about how many percentage of the web page already downloaded and how many hours needed to complete this task. Another important aspect is information fit-to-task, information presented on a website is accurate and appropriate for the task at hand (Loiacono et al., 2007). Good architecture is fundamental to deal a website's requirements, to ensure structural scalability, flexibility, security and to fulfil performance demands for this time and in the future. A completed site should comply with acknowledged programming standards. As the web keep on growing as a competitive tool for business applications, there is a need to comprehend the relationship between business performance and web usability. Much of the previous research has discussed the website development from a set of usability factors (Green and Pearson, 2006; Seffah et al., 2006). When we applied accessibility test online to examine whether the web portals have accessibility errors on their respective web pages, 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 second priority, then users will have some problems to access the web, while if the website already satisfied the third criteria, then user 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 in order 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 pairwise

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 a three objective optimisation problem in order 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 on the website evaluation, most studies 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 for six university sites from different countries tested using this factor (Olsina et al., 2001). Website architecture is classified into content and design (Huizingh, 2000), and each category 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 perspective. Best websites selected by experts and users are investigated in order to identify the common characteristics of them (Ivory and Hearst, 2002; Sinha et al., 2001). To empirically determine whether content is more important than graphics, Webby Award 2000 data set is examined to differentiate the factors of best websites from the factors of other websites (Sinha et al., 2001). Webby Award evaluators use five specific criteria and the general experience. 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

A comprehensive review on the automation of user interface usability evaluation is discussed in literature including automated website evaluation tools (Ivory and Hearst, 2001). In this survey, the usability evaluation method is summarised 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 to the website usability evaluation (Brajnik, 2000). The comparison of automated

evaluation tools using consistency, adequate feedback, situational navigation, efficient navigation and flexibility as the characteristic of usability is explored in this research (Brajnik, 2000). Website evaluation model based on the stages of a transaction in the e-market is another approach (Schubert and Selz, 1999). There are three stages of the e-commerce information stage, contract stage and payment stage and assume that the communication stage in the cyber community plays an important role (Schubert and Selz, 1999). Their website evaluation model is based on the stages and was applied to the Swissair website. 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 and applies it to the UsableNet LIFT, an automated website evaluation tool (Brajnik, 2000, 2002). The validity of a set of website evaluation criteria is verified using the Webby Award 2000 data set (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 (TQM), information technology (IT) and web engineering (Sudhahar et al., 2009). The quality of service (QoS) in the internet and richer understanding of internet accounting taxonomy, such as attributes, parameters, protocols, records and metering tools need to be updated or replaced (Hak Ju, 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 is 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 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's 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 include 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 Guidelines (WCAG) or Section 508 Guidelines. In 1998, US government, the federal law Rehabilitation Act 508, requires all e-information technologies allow handicap people to use them. Therefore, every website is required to provide accessibility to all and this guideline becomes an evaluation criterion of automated website evaluation tools. Max of 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.

Today, usability is recognised as a major quality and success factor of websites. A wide range of usability evaluation techniques has 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 became much desired (Brajnik, 2000; Cooper, 2008; Ivory and Hearst, 2001) because they required usability specialists to conduct them or to analyse evaluation results, which is very resource consuming especially for very large, 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 it 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 WCAG (http://www.w3c.org/TR/WCAG10) and Section 508 Guidelines (http://www. Section 508.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 (Lin 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 web page design has their own characteristics and this characteristic has drawbacks and benefits. There is a mechanism for measuring the effects of the web page 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 main factor that will influences 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 website to be a good quality website. Tested factors consist of average server response time, number of

component per page, web page loading time and web page size in byte. A standard international download time for this performance can be used as a reference to categorise the tested web page. 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. For leveraging the effectiveness of the continuous quality improvement, developer community has been aggressive in attaining TQM strategies by implementing ISO 9001:2000 standard (Sakthivel et al., 2007).

Broken links can give bad impact for the credibility of a website. Credibility is especially 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
Web page loading time	<30 sec
Web page size in byte	<64 kB

Source: Amerson et al. (2001).

3   Methodology

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

1   to propose the new methodology for evaluating the quality of airlines website

2   to determine the best airlines 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 national airlines portals of a selected number of countries in Asia: Singapore Airlines, Korean Airlines, Japan Airlines, Cathay Pacific and Malaysia Airlines. This data of quality website from airlines website will be taken more than 30 trails on various occasions on the different period of time. This data have been taken from 29 March 2009 until 20 May 2009. Using website diagnostic tools and four selected methods (linear weightage model (LWM), AHP, fuzzy analytical hierarchy process (FAHP) and new hybrid model (NHM)) the aim of this research will be explored. Data were analysed by using non-parametric statistical test. To analyse whether there is 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 web page speed analyser online service (http://www.websiteoptimization.com). List of performance measured and reported by this service includes total size, number of objects (HTML, images, CSS and scripts) and download times on a 56.6 kBps connection, another available web page online tools that we used for testing quality is http://validator.w3.org/checklink, which was utilised in order to monitor broken links in the HTML code of the portals, while 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 W3C, the main international standards organisation for the World Wide Web. A website tool for measuring link popularity website (www.linkpopularity.com) is used to determine the amount and quality of links that are made to a single website from many websites, this based on the page-rank analysis.

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

3.2 Sample data

To get the data for this research, we examined airlines websites from five Asian countries: the airlines websites were not randomly selected, but a careful process was undertaken. Rather than selecting any generic airlines websites this research attempted to evaluate the websites that are considered to be leaders in the area IT implementation based on result of a survey conducted by pingdom and Skytrax Company. By doing such an approach, it was felt that measures of 'best practices' could emerge. As explained before, we examined the national airlines portals of a selected number of countries and their web addresses are provided along with the names, which are Singapore (http://www.singaporeair.com), Korean (http://www.koreanair.com), Japan (http://www.jal.com), Hong Kong (http://www.cathaypacific.com) and Malaysia (http://www.malaysiaairlines.com). This data in Table 7 will be taken more than 30 trails on various occasions on the different period of time.

3.3 Linear weightage model

This model is very easy and mostly depending upon decision-maker's 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 assigned weight to each individual criterion in order to determine the relative importance of each one. These weights play a vital role in decision-making process and extremely affect the final decision. First of all decision-maker has to identify all criteria that involve in the certain process before performing any other steps. After identifying all the criteria related to website selection decision, decision-maker has to determine threshold for each criterion. In fact, threshold can be divided into two types, i.e. maximum and minimum.

To establish a threshold to criterion, decision-maker 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 categories as 'smaller is better' and the threshold for this type of criteria must be maximum. On the other hand, others criteria can be considered as 'larger is better' such as traffic, page rank, frequency of update and design optimisation where thresholds must be minimum.

3.4 Analytic hierarchy process

Analytic hierarchy process (AHP) was originally designed by Saaty (1980) to solve complicated multi-criteria decision problem, beside that AHP is 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 airlines website. AHP is a popular model to aggregate multiple criteria for decision-making (Yuen and Lau, 2008). In AHP, the problems are usually presented in a hierarchical structure and the decision-maker is guided throughout a subsequent series of pairwise 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. In the constantly fluctuating of the website, evaluation cannot survive without comprehensive quality factor identification and evaluation. The AHP methodology is demonstrated by applying it to the quality factors with the airlines 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 airlines 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.

3.5 Fuzzy analytic hierarchy process

In 1965, Lotfi A. Zadeh introduced a new approach to a precise theory of approximation and vagueness based on generalisation of standard set theory to fuzzy sets. 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. Their 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 as 1. Values between 0 and 1 represent intermediate degrees of membership. Weight parameter for AHP and FAHP depicts in Table 3.

Decimal judgements, such as 3.5, are allowed for fine tuning, and judgements greater than 9 may be entered, though it is suggested that they be avoided.
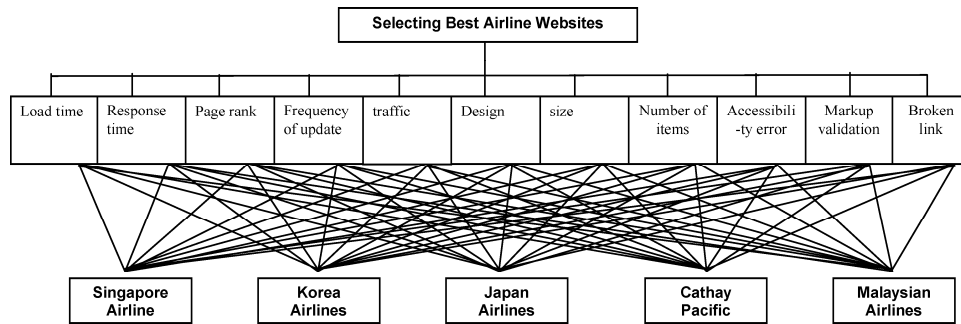
Figure 1 AHP/FAHP model of airlines websites

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

Linguistic expressions	Fuzzy AHP			AHP
	a_1	a_2	a_3	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

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

Data were analysed by using non-parametric statistical test. 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. To answer this problem, we use 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 pairwise comparison (*t*-test), but the critical value is different.

4   Result and discussion

First column in Table 4 shows the criteria of the quality website. Criteria involves 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 country airlines website performance value.

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 show that most of the airlines websites in Asian cannot meet the criteria as a high-quality website. Most of server response, load times, size and number of items exceed the value standardised by IBM, except Malaysia airlines websites in load time, size and number of items criteria. Implementation of the W3C's HTML validator highlighted that only Japan Airlines of the Asian airlines website had HTML 4.01 valid entry page, most of it did not have DOCTYPE declarations. Consequences of this problem will be on the portability and development of the website. In term of broken link, two airlines website or 40% of the sample have a broken link.

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

$$W_{s\,max} = \frac{Max - Website}{Max - Min} \tag{1}$$

$$W_{s\,min} = \frac{Website - Min}{Max - Min} \tag{2}$$

where

W_{smax} = specific website value that has maximum type of threshold with respect to a particular attribute/criterion

W_{smin} = 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 4 Testing result for websites performance based on criteria

Website/ criteria	Measuring unit	SIA	KAL	JAL	Cathay	MAS
A	Second	91.91	5.16	35.50	42.23	0.32
B	Second	1.35	1.92	1.56	1.10	1.52
C	Number link	1,180.00	919.00	326.00	1,310.00	765.00
D	Update during two month	60.00	60.00	60.00	60.00	60.00
E	Number	971,100.00	533,000.00	410,400.00	868,200.00	861,500.00
F	Percentage	25.00	27.00	61.00	92.00	89.00
G	kB	408,003.00	21,865.00	123,919.00	145,666.00	582.00
H	Number	53.00	4.00	54.00	66.00	1.00
I	Number of error	2.00	12.00	0.00	26.00	0.00
J	Number of error	141.00	25.00	0.00	444.00	1.00
K	Number of broken link	2.00	0.00	0.00	28.00	0.00

Table 5 Final result for airlines website (LWM)

Website/criteria	SIA	KAL	JAL	Cathay	MAS	Weight
A (load time)	0.000	0.947	0.616	0.542	1.000	0.16
B (response time)	0.693	0.000	0.437	1.000	0.490	0.14
C (page rank)	0.868	0.603	0.000	1.000	0.446	0.12
D (frequency of update)	1.000	1.000	1.000	1.000	1.000	0.11
E (traffic)	1.000	0.219	0.000	0.816	0.805	0.11
F (design optimisation)	0.000	0.030	0.537	1.000	0.955	0.11
G (size)	0.000	0.948	0.697	0.644	1.000	0.09
H (number of items)	0.200	0.954	0.185	0.000	1.000	0.07
I (accessibility error)	0.923	0.538	1.000	0.000	1.000	0.05
J (markup validation)	0.682	0.944	1.000	0.000	0.998	0.04
K (broken link)	0.93	1.00	1.00	0.00	1.00	0.02
Sum	0.52	0.58	0.50	0.70	0.84	
Rank	4	3	5	2	1	

The idea of using formulae (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 concepts which represented in formulae (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 over all.

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

We give every criteria with the appropriate weight depend on their 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). Then get 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 is regarding 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, Malaysia Airlines website has the highest score of 0.84 in comparison with the rest of airlines websites. As a result, the proposed LWM model rank for airlines website is: Malaysia Airlines (score: 0.84), Cathay Pacific Airlines (score: 0.70), Korea Airlines (score: 0.58), Singapore Airlines (score: 0.52) and the last rank is Japan Airlines (score: 0.50). Table 6 presents the weights of airlines website associated with each of the website quality criteria based on evaluation of their contribution towards overall quality using AHP model.

The final score obtain for each website across each criterion is calculated by multiplying the weight of each criterion with the weight of each website. Website which has got 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 pairwise comparison approach. Fundamental scale for pairwise comparisons developed to solve this problem (Saaty, 1980). The pairwise comparison matrix A, in which the element a_{ij} of the matrix is the relative importance of the ith factor with respect to the jth factor, could be calculated as

$$A = \begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix} \tag{4}$$

4   There are $n(n-1)/$ judgements required for developing the set of matrices in step 3. Reciprocals are automatically assigned to each pairwise comparison, where n is the matrix size. Table 7 depicts the final scores of websites based on AHP evaluation method.

In accordance with the results generated by the proposed model, Malaysia Airlines website has the highest score of 0.269 in comparison with the rest of airlines websites. As a result, the proposed AHP model rank for airlines website is: Malaysia Airlines (score: 0.269), Cathay Pacific Airlines (score: 0.252), Korea Airlines (score: 0.177), Singapore Airlines (score: 0.169) and the last rank is Japan Airlines (score: 0.134). Table 8 presents the weights of airlines website associated with each of the website quality criteria, based on evaluation of their contribution towards overall quality using 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.

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

$$\mu\left(\frac{x}{\tilde{M}}\right) = \begin{cases} 0, & x < 1 \\ (x-l)/(m-l), & l \leq x \leq m \\ (u-x)/(u-m), & m \leq x \leq u \\ 0, & x > u \end{cases} \qquad (5)$$

Table 6   Weight of criteria and website AHP

Website/criteria	SIA	KAL	JAL	Cathay	MAS	Weight
A (load time)	0.030	0.286	0.157	0.095	0.433	0.270
B (response time)	0.259	0.058	0.110	0.413	0.159	0.197
C (page rank)	0.253	0.136	0.033	0.506	0.071	0.148
D (frequency of update)	0.200	0.200	0.200	0.200	0.200	0.107
E (traffic)	0.418	0.101	0.055	0.253	0.172	0.076
F (design optimisation)	0.040	0.058	0.183	0.410	0.309	0.052
G (size)	0.074	0.285	0.117	0.085	0.439	0.042
H (number of items)	0.085	0.308	0.064	0.035	0.509	0.042
I (accessibility error)	0.211	0.116	0.313	0.048	0.313	0.030
J (markup validation)	0.069	0.172	0.416	0.029	0.313	0.021
K (broken link)	0.147	0.272	0.272	0.036	0.272	0.016

Table 7   Final result evaluation (AHP)

Website/criteria	SIA	KAL	JAL	Cathay	MAS
A (load time)	0.008	0.077	0.042	0.026	0.117
B (response time)	0.051	0.011	0.022	0.081	0.031
C (page rank)	0.037	0.020	0.005	0.075	0.011
D (frequency of update)	0.021	0.021	0.021	0.021	0.021
E (traffic)	0.032	0.008	0.004	0.019	0.013
F (design optimisation)	0.002	0.003	0.010	0.021	0.016
G (size)	0.003	0.012	0.005	0.004	0.018
H (number of items)	0.004	0.013	0.003	0.001	0.021
I (accessibility error)	0.006	0.003	0.009	0.001	0.009
J (markup validation)	0.001	0.004	0.009	0.001	0.007
K (broken link)	0.002	0.004	0.004	0.001	0.004
Sum	0.169	0.177	0.134	0.252	0.269
Rank	*4*	*3*	*5*	*2*	*1*

Table 8   Weight criteria and website (FAHP)

Website quality criteria	SIA	KAL	JAL	Cathay	MAS	Weight
A (load time)	0.000	0.364	0.107	0.000	0.529	0.377
B (response time)	0.284	0.000	0.167	0.401	0.148	0.291
C (page rank)	0.276	0.000	0.000	0.724	0.000	0.216
D (frequency of update)	0.200	0.200	0.200	0.200	0.200	0.114
E (traffic)	0.498	0.000	0.000	0.316	0.186	0.003
F (design optimisation)	0.000.	0.000	0.207	0.417	0.376	0.000
G (size)	0.000	0.428	0.000	0.000	0.572	0.000
H (number of items)	0.000	0.389	0.000	0.000	0.611	0.000
I (accessibility error)	0.264	0.103	0.316	0.000	0.316	0.000
J (markup validation)	0.000	0.132	0.460	0.000	0.408	0.000
K (broken link)	0.144	0.285	0.285	0.000	0.285	0.000

Figure 2   A triangular fuzzy number, $\tilde{M}$
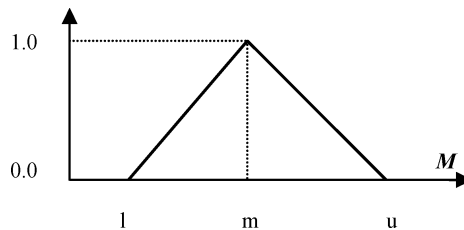
In applications, it is easy to work with TFNs because of their simple computation, and they are useful in promoting representation and information processing in a fuzzy environment. In this research implementation of TFNs in the FAHP is adopted. We have to deal with fuzzy numbers when we want to use 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 $X = \{x_1, x_2, x_3, \ldots, x_n\}$ be 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:

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

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 ith object is defined as

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

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 $M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$ 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:

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

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

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

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

Assume that $d(A_i) = \min V(S_i \geq S_k)$ 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 illustrates Equation (9) where d is the ordinate of the highest intersection point D between μ_{M_1} and μ_{M_2} to compare M_1 and M_2, we need both the values of $V\left(M_1 \geq M_2\right)$ and $V\left(M_2 \geq M_1\right)$

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 airlines website associated with each of the website quality criteria, based on evaluation their contribution towards overall quality using FAHP model.

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

The website which has the highest score is suggested as the best website for the proposed FAHP model. Cathay Pacific website has the highest score of 0.29682 in comparison with the rest of airlines websites. As a result, the proposed FAHP model rank for airlines website is: Cathay Pacific Airlines (score: 0.29682), Malaysia Airlines (score: 026586), Singapore Airlines (score: 0.16655), Korea Airlines (score: 0.16003) and the last rank is Japan Airlines (score: 0.11174). Table 10 depicts the scores of websites using hybrid model evaluation. The most important thing is regarding the final results, the website which has the highest score is suggested as the best website for the proposed hybrid model.

Table 10 derived from Table 5 and Table 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.

Table 11 depicts the final scores of websites. The most important thing is regarding the final results, the website which has the highest score is suggested as the best website for the proposed hybrid model. Malaysia Airlines website has the highest score of 0.73 in comparison with the rest of airlines websites. As a result, the proposed hybrid model rank for airlines website is: Cathay Pacific (score: 0.83), Malaysia Airlines (score: 0.73), Korea Airlines (score: 0.60), Singapore Airlines (score: 0.51), and the last rank is Japan Airlines (score: 0.47).

Table 12 depicts the final ranking of airlines websites based on four specific methods. Malaysia Airlines website has the highest in comparison with the rest of airlines websites for LWM and AHP methods. As a result, the evaluation model ranking (LWM and AHP) for airlines website is: Malaysian Airlines, Cathay Pacific, Korea Airlines, Singapore Airlines, and Japan Airlines. Inconsistency occurred for the FAHP and hybrid model, because the ranking list for FAHP method is: Cathay Pacific, Malaysia Airlines, Singapore Airlines, Korea Airlines and Japan Airlines, while using hybrid model the ranking list are: Cathay Pacific, Malaysia Airlines, Korea Airlines, Singapore Airlines and Japan Airlines.

To analyse whether there is differences among the ranking composition methods (Table 13), we ran 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 check the ranking, note that the sum of the four rank sums is $18 + 9 + 6 + 17 = 50$, 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

$$\sum \text{SR}_i = \frac{rc(c+1)}{2} = \frac{5(4)(5)}{2} = 50.$$

Now compute the Friedman statistic

$$x_F^2 = \left[\frac{12}{rc(c+1)} \sum_i \left(\text{SR}_i^2 \right) \right] - 3r(c+1)$$

$$x_F^2 = 12.6$$

Figure 3 The intersection between M_1 and M_2
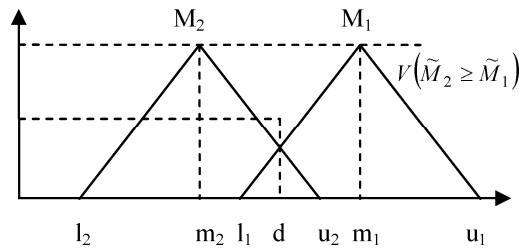
Table 9 Final result FAHP

Website/criteria	SIA	KAL	JAL	Cathay	MAS
A (load time)	0.00000	0.13723	0.04034	0.00000	0.19943
B (response time)	0.08264	0.00000	0.04860	0.11669	0.04307
C (page rank)	0.05962	0.00000	0.00000	0.15638	0.00000
D (frequency of update)	0.02280	0.02280	0.02280	0.02280	0.02280
E (traffic)	0.00149	0.00000	0.00000	0.00095	0.00056
F (design optimisation)	0.00000	0.00000	0.00000	0.00000	0.00000
G (size)	0.00000	0.00000	0.00000	0.00000	0.00000
H (number of items)	0.00000	0.00000	0.00000	0.00000	0.00000
I (accessibility error)	0.00000	0.00000	0.00000	0.00000	0.00000
J (markup validation)	0.00000	0.00000	0.00000	0.00000	0.00000
K (broken link)	0.00000	0.00000	0.00000	0.00000	0.00000
Total	*0.16655*	*0.16003*	*0.11174*	*0.29682*	*0.26586*
Rank	*3*	*4*	*5*	*1*	*2*

Table 10 Maximum minimum criteria (hybrid)

Website/criteria		SIA	KAL	JAL	Cathay	MAS	Weight
A (load time)	Max	0.000	0.947	0.616	0.542	1.000	*0.377*
B (response time)	Max	0.693	0.000	0.437	1.000	0.490	*0.291*
C (page rank)	Min	0.868	0.603	0.000	1.000	0.446	*0.216*
D (frequency of update)	Min	1.000	1.000	1.000	1.000	1.000	*0.114*
E (traffic)	Min	1.000	0.219	0.000	0.816	0.805	*0.003*
F (design optimisation)	Min	0.000	0.030	0.537	1.000	0.955	*0.000*
G (size)	Max	0.000	0.948	0.697	0.644	1.000	*0.000*
H (number of items)	Max	0.200	0.954	0.185	0.000	1.000	*0.000*
I (accessibility error)	Max	0.923	0.538	1.000	0.000	1.000	*0.000*
J (markup validation)	Max	0.682	0.944	1.000	0.000	0.998	*0.000*
K (broken link)	Max	0.93	1.00	1.00	0.00	1.00	*0.000*

Table 11   Final result

Website quality criteria	SIA	KAL	JAL	Cathay	MAS
A (load time)	0.00	0.36	0.23	0.20	0.38
B (response time)	0.20	0.00	0.13	0.29	0.14
C (page rank)	0.19	0.13	0.00	0.22	0.10
D (frequency of update)	0.11	0.11	0.11	0.11	0.11
E (traffic)	0.00	0.00	0.00	0.00	0.00
F (design optimisation)	0.00	0.00	0.00	0.00	0.00
G (size)	0.00	0.00	0.00	0.00	0.00
H (number of items)	0.00	0.00	0.00	0.00	0.00
I (accessibility error)	0.00	0.00	0.00	0.00	0.00
J (markup validation)	0.00	0.00	0.00	0.00	0.00
K (broken link)	0.00	0.00	0.00	0.00	0.00
Total	0.51	0.60	0.47	0.83	0.73
Rank	*4*	*3*	*5*	*1*	*2*

Table 12   Final result for airlines websites performance

Method	SIA	KAL	JAL	Cathay	MAS
LWM	0.52 (4)	0.58 (3)	0.50 (5)	0.70 (2)	0.84 (1)
AHP	0.169 (4)	0.177 (3)	0.134 (5)	0.252 (2)	0.269 (1)
FAHP	0.166 (3)	0.160 (4)	0.112 (5)	0.297 (1)	0.266 (2)
Hybrid	0.51 (4)	0.60 (3)	0.47 (5)	0.83 (1)	0.73 (2)

Table 13   Airlines websites ranking based on method

	Original data					Ranked data			
	LWM	*AHP*	*FAHP*	*Hybrid*		*LWM*	*AHP*	*FAHP*	*Hybrid*
SIA	0.52	0.169	0.16655	0.51	SIA	4	2	1	3
KAL	0.58	0.177	0.16003	0.60	KAL	3	2	1	4
JAL	0.5	0.134	0.11174	0.47	JAL	4	2	1	3
CATHAY	0.7	0.252	0.29682	0.83	CATHAY	3	1	2	4
MAS	0.84	0.269	0.26586	0.73	MAS	4	2	1	3
					Sri	18	9	6	17

If we find the place on the Friedman Table for four columns and five rows, we find that the *p*-value for $x_F^2 = 12.6$ is 0.0056. Since the *p*-value is below $\alpha = 0.05$, reject the null hypothesis. Since the computed FR 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. You 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/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.

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) (Table 14). The Bonferroni *t*-test is the same as a normal pairwise comparison (*t*-test), but the critical value is different. Because we are allowed to make many comparisons, we have to control for family wise error by reducing the per comparison level. The overall level will be set to 0.05, and the individual per comparison levels will be equal to 0.05 divided by the total number of possible comparisons. We can make a total of $4C_2 = 6$ different pairwise comparisons between the four means. In practice, we cannot do all of these comparisons, but remember that we will always have to set the error rate according the total number of possible comparisons.

Step 1 Calculate the t' statistics.

General formula:

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

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

$$SS_{Treatment} = \sum_{i=1}^{4} \frac{y_i^2}{n} - \frac{y^2}{n}$$
$$= 21$$

$$SS_E = SS_T - SS_{Treatment}$$
$$= 25 - 21 = 4$$

$$MS_{Treatment} = \frac{SS_{Treatment}}{a-1}$$
$$= \frac{21}{3} = 7$$

$$MS_E = \frac{SS_e}{[a(n-1)]}$$
$$= \frac{4}{[4(4)]} = 0.25$$

General formula:

$$t^1 = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\left(MS_{error} / n\right) + \left(MS_{error} / n\right)}} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\left(2\left(MS_{error}\right)\right) / n}}$$

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 α will be:

$$\alpha = \frac{\alpha_{FW}}{k}$$
$$= \frac{0.05}{4} = 0.0125$$
$$df = df\ MS_{error} = 16$$

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′* = 5.69 (significant), LWM vs. FAHP: *t′* = 7.59 (significant), LWM vs. Hybrid: *t′* = 0.633 (not significant), AHP vs. FAHP: *t′* = 1.89 (not significant), AHP vs. Hybrid: *t′* = −5.06 (not significant), FAHP vs. Hybrid: *t′* = −6.96 (significant). Therefore, we can conclude that different methods are significantly different.

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*
$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{(2(MS_{error})) / n}}$	$\dfrac{\overline{x}_1 - \overline{x}_3}{\sqrt{(2(MS_{error})) / n}}$	$\dfrac{\overline{x}_1 - \overline{x}_4}{\sqrt{(2(MS_{error})) / n}}$	$\dfrac{\overline{x}_2 - \overline{x}_3}{\sqrt{(2(MS_{error})) / n}}$	$\dfrac{\overline{x}_2 - \overline{x}_4}{\sqrt{(2(MS_{error})) / n}}$	$\dfrac{\overline{x}_3 - \overline{x}_4}{\sqrt{(2(MS_{error})) / n}}$
$\dfrac{3.6 - 1.8}{\sqrt{(2(0.25)) / 5}}$	$\dfrac{3.6 - 1.2}{\sqrt{(2(0.25)) / 5}}$	$\dfrac{3.6 - 3.4}{\sqrt{(2(0.25)) / 5}}$	$\dfrac{1.8 - 1.2}{\sqrt{(2(0.25)) / 5}}$	$\dfrac{1.8 - 3.4}{\sqrt{(2(0.25)) / 5}}$	$\dfrac{1.2 - 3.4}{\sqrt{(2(0.25)) / 5}}$
= 5.69	= 7.59	= 0.633	= 1.89	= −5.06	= −6.96
Significant	Significant	Not significant	Not significant	Not significant	Significant

5   Conclusion

In this paper, we evaluate the quality of airlines websites with the sample of five Asian carrier flag airlines. Using a series of online diagnostic tolls, we examined many dimensions of quality, and each dimension will be measured by specific test online. The result of this study confirmed that the website presence of airlines website is neglecting performance and quality criteria. It is clear in our research that more effort is required to meet with these criteria in the context of website design. This suggests that web developer responsible for airlines website should follow and encourage the use of recognised guidelines when designing website. We propose a methodology for determining and evaluate the best airlines websites based on many criteria of website quality, consist of LWM, AHP, FAHP and one NHM. By applying hybrid model between

FAHP and LWM 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. 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 website.

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 and 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 Co, Pte. Ltd.

Bouch, A., Kuchnisky, A. and Bhatti, N. (2000) 'Quality is in the eye of the beholder: meeting users' requirements for Internet quality of service', *Human Factors in Computing Systems*. The Hague, Netherlands: ACM.

Brajnik, G. (2000) 'Automatic web usability evaluation: What needs to be done?', *Proceedings of 6th Conference on Human Factors and the Web*, Available at: www.tri.sbc.com/nfweb/brajnik/hfweb-brajnik.html.

Brajnik, G. (2002) 'Quality models based on automatic web testing', 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.

Cooper, M. (2008) *W3C Web Content Accessibility Guideline* (WCAG) 2.0.

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.

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.

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', *IEEE International Fuzzy Systems Conference Proceeding*, Seoul, South Korea, Vol. 2, pp.726–731.

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.

Hak Ju, K. (2009) 'New accounting system for the QoS-enabled multiservice internet: technical overview', *Int. J. Business Information Systems,* Vol. 4, pp.173–194.

Huizingh, E. (2000) 'The content and design of web sites: an empirical study', *Information* and *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', *The SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves*. Minneapolis, Minnesota, USA: ACM.

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

Josep, D., Ana, P., Julio, S. and Jose, G. (2007) 'A user-focused evaluation of web prefetching algorithms', *Computer Communications*, Vol. 30, pp.2213–2224.

Krishnamurthy, B. and Wills, C. (2006) 'Cat and mouse: content delivery tradeoffs in web access', *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* and *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* and *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 Systems with Applications*, Vol. 36, pp.8218–8226.

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

Mcinerney, C. (2000) 'Establishing and maintaining trust in online systems', in M. Williams (Ed.), *The 21st National Online Meeting Information Today*. Medford, N.J.

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 Systems with Applications*, Vol. 36, pp.190–198.

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*, pp.1–21.

Page, L. and Brin, S. (1998) 'The anatomy of a large-scale hyper textual web search engine', *Computer Networks and ISDN Systems*, Vol. 30, Nos. 1–7, Article FP 11.

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.

Sakthivel, M., Devadasan, S.R., Vinodh, S., Ramesh, A. and Shyamsundar, S. (2007) 'ISO 9001:2000 based quality information management responsibility system', *Int. J. Business Information Systems*, Vol. 2, pp.217–237.

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', *Proceeding of the 4th Conference on Human Factors and the Web, 1998*, 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', *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* and *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.

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

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) 'Optimizing 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.