

Role of Web Search Engines in Knowledge Acquisition

¹ Mohd Shukri Ab Yajid*, ¹S.M. Ferdous Azam, ¹Sakinah

¹Management and Science University

*Corresponding author: shukri@msu.edu.my

ABSTRACT

The purpose of the current analysis is to explore the usage of web engines for the sake of knowledge gaining and sharing. First this study used the appropriate research tools; applications and other related peripherals. Next mentioned are the research procedure elaboration and data sub-categorization explanation. Statistical analysis is applied to study the distribution of download speed and the relationship between the sizes of websites with download speed by applying the simple regression method.

Keywords: web search engine. Knowledge gaining, Malaysia

Correspondence:

Mohd Shukri Ab Yajid
Management and Science University
Corresponding author: shukri@msu.edu.my

INTRODUCTION

Malaysia embarked on its journey to be a developed nation by the 2020 when the master plan was laid by the former Prime Minister of Malaysia, Dr. Mahathir Mohammad. The master plan aims to create a developed nation with secular social values based on its unique cultural diversity and unity. In achieving this goal, the Multimedia Super Corridor (MSC) was created to develop and penetrate the world technology market. This is crucial as technology is being applied in every industry worldwide. The Multimedia Super Corridor (MSC) was launched in 1996 to support and promote Malaysia Information Technology (IT) industry parallel with global technological demand and to position the country as a major Information Communication Technology (ICT) and multimedia hub. In 2005, the MSC was extended to Bayan Lepas and Kulim Technological Park in the southern and northern regions of Peninsular Malaysia. This demonstrates the growing role of MSC and the demand for (ICT) infrastructure in Malaysia. The MSC is connected with world class infrastructure - a 10 gigabit digital fiber optic backbone and high-end global communication system (Malinský & Jelínek, 2010; De Silva et al., 2018a; De Silva et al., 2018b; Nikhashemi et al., 2013).

In order for a company to be granted the MSC status, it is required by the MDC that it fulfills the criteria as well as being successful in its application. MSC status companies are entitled to incentives from the Malaysian government guaranteed by the 10 Bill of Guarantees. Companies are entitled to ten years of tax relief or 100% investment tax allowance besides other benefits such as unrestricted employment of foreign knowledge workers and no duty import on multimedia equipment. These incentives are meant to aid and promote the local companies to establish a firm foothold in the industry and to attract foreign investment into the country at the same time. As illustrated in Figure 1, as of December 6, 2004, there is 1153 MSC status companies, of which 71% are Malaysian-owned, 26.4% are foreign-owned and 2.6% are joint ventures. The number of MSC status companies have risen to 18.5% in 2004 which consists of 184 companies (MCMC 2004). The government's efforts in promoting the ICT

industry and the MSC as the flagship has proven the importance and potential of the ICT sector as a contributor to the country's GDP.

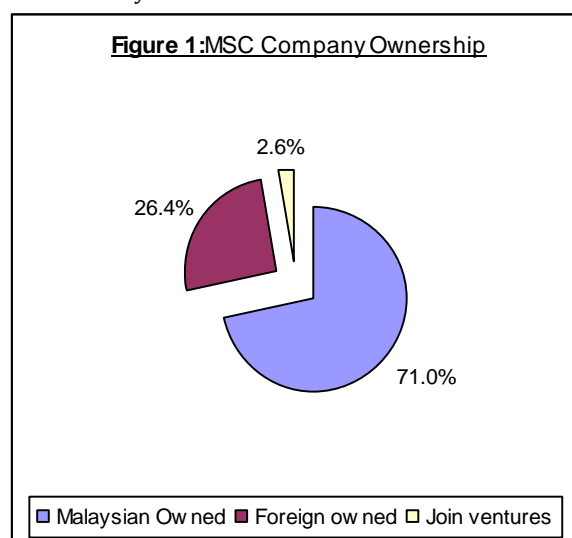


Figure 1: MSC Company Ownership

Growth of MSC status companies has increased 19% from the year 2003 to 2004, where the highest increase is from Software Development - Business Applications sector which recorded 35.9% increase. However, there is a drop in Shared Service sector; negative 25.9%.

Measuring the efficiency of company websites is crucial in view that a web presence is the key component in the marketing strategy for technology based companies. Efficiency of the information intended to be conveyed to the audience is studied with predetermined characteristics. Measuring the website efficiency is vital since there is no yardstick for the effectiveness of information published in cyberspace. Therefore, the aim of this research is to investigate the efficiency of MSC Status Companies websites. It is important to understand the significant role of a website. It requires resources to be developed and the content will reflect not just the entity

that it represents but indirectly the MSC as well as the web pages of MSC status companies are analyzed with these objective:

1. To determine the availability of descriptive factor in the company's web page. Descriptive factors consist of:

- a. Type of top-level domain name
- b. Information of authorship
- c. Availability of keyword
- d. Availability of description
- e. Currency which is derived from the last updated information.

2. Measuring the technical factors of the web page from:

a. The size of a web page. The null hypothesis for this test is:

H0: Website loads in equal or less than 20 seconds.

H1: Website does not load in equal or less than 20 seconds.

b. The time taken to load the web page. The null hypothesis for this test is:

H0: Website size in equal or less than 20 kilobytes.

H1: Website size is more than 20 kilobytes.

c. The technical performance of Malaysia's website (MSC's status company websites) against websites from the United States, UK & Ireland and Japan. The null hypothesis for this test is:

H0: $\mu_1 = \mu_2$

H1: $\mu_1 \neq \mu_2$

Where is the mean for websites from the United States, UK & Ireland and Japan.

LITERATURE REVIEW

In this chapter, related research and articles are reviewed and related elements of research are synthesized to support the criteria of this study. The journal articles are mostly acquired online through Emerald database accessed from Siti Hasmah digital library. Day 1997 purposes a model which organization can use to monitor website effectiveness. In general, effectiveness of a website depends on developers understanding on who the customer is and what the customer wants to see and what the customer perceives as quality. Measuring effectiveness of websites based on the number of hits is inadequate. The right measurement is to get the number of the "right hits"; people that visit the site and do what the developer wants for them to do. Websites should be structured in a manner where visitors have the opportunity to view the entire content before browsing further. Web pages should have governing rules that project integrity and credibility of the organization it represents. Web developers must deliver what they promise, requests for additional information should be attended to immediately. The arbitrary reason why we go online is speed, therefore access time should be a priority to any web developer in creating a site. It is agreeable that measuring the effectiveness of a website is a daunting task and feedback might be the only accurate measure for it. In order to get feedback, a sample group should be created consists of invited users with an appropriate reward system. In order to collect specific information, questionnaires should be specific and less open-ended (Chau, Zeng, Chen, Huang, & Hendriawan, 2003).

The tolerable waiting time (TWT) for an average user is 2 to 4 seconds according to a study conducted by Nah F. (2004). The study has concluded that an average user will only wait for two seconds to retrieve simple information

from a website in order to maintain continuity of human thought processing. Even though the waiting time may vary due to its importance and indication of download progress, browsers or web pages that have no indication of download progress have only 15 seconds of TWT form (from?) the user.

Factors that influence TWT is mostly behavioral rather than technical but technical factors do influence behavior. Nah has a distinct two behavior category in accessing the Internet; netsurfing and querying. Netsurfing is a scenario where the user is surfing various websites without any goal; this behavior is more tolerant to longer waiting periods compared to queries pertaining to a specific topic.. Querying has a clear purpose and objective in accessing the Internet, these users have lower TWT. User TWT increases by technical factors such as connection speed, when the user is aware of slow connection speed, TWT will increase and with higher speed the TWT will decrease. Content does affect TWT, user who knows multimedia content in a webpage will wait longer for loading as it does increase the experience visiting the site (Eirinaki & Vazirgiannis, 2003).

Past studies how consumers gather information and the potential. With the creation of e-market, consumers have the ability to gather information more efficiently and with minimal imperfect information, it also increases the chances of buyer and seller matching criteria. Searching online costs less than a brick-and-mortar market as it reduces the cost of communicating information. There is a huge amount of dynamic information that is available online which is useful for user queries in decision making. There are three perceived antecedents benefit of online search; ease of use of technology, effectiveness of online search and user satisfaction. These three perceived benefits are also the main factor online query is highly used for information gathering besides the fact that it minimizes cost and hassle. In order to retrieve the right information, it is first required to obtain the right site, users generally use online search engine by entering the key criteria. Common search criteria is the keyword. This is contained in the description of the website which is available if provided by the site developer. However, if it is not available, the site can be missed by the search engine (Ma, Li, & Yang, 2016).

Li 2002 analyzed five major online versions of dailies to illustrate different aspects of web design that contributes to the efficiency of information retrieval. Efficiency of information retrieval is measured by the options that is available for news readers to select news items, processes to access news stories, required time to acquire the news content and the amount of information that was retrieved in a specific time frame. The highest retrieval efficiency score was earned by the daily that has offered the most choices or highest number of information in the least retrieval time. Even though loading speed is critical to measure efficiency, but flow of information through appropriate web design, a web page that is unable to retrieve information effectively might experience design flaw. The site can be visually appealing but the information available can be useless to readers. This affects the efficiency of information conveyed by the news publisher. Understanding the need and requirements of users in obtaining information from the Internet has shown that there are many factors that affect retrieval efficiency, even though speed is the primary criterion of a web page, information relevance, presentation and flow is equally critical (Cheng, 2009).

The question to purchase off-the-shelf or custom built solution requires several factors to be considered. Large organization may prefer custom built CM solutions as they have more IT personnel and requires extensive customized system. On the other hand, additional IT investment may be necessary for implementation and IT staff might be distracted from priority projects. Solutions that are complex faces user difficulty to adapt, additional training is required to resolve this issue and most of the time incur cost. It is recommended that CM solution contains these features; versatile server application, content authoring, version control, easy browser access, user friendly interface, authentication and group creation. As the demand for CM increases by small and mid-size organization, CM will be more affordable and attractive. Past studies identified deferent functions and features for web monitoring systems. WebMon, a monitoring system was developed as a result of this study. Factors that were considered are; text, links, image, layout and last updated date. Study observation shows that out of 105 test sample pages, 96.1% contains hyperlinks, 78.6% contains images and 43.6% contains last updated date. It is identified that Java is still the main programming language used for web applications as it can easily be integrated into many common browsers. Business domains are highly active; it is being updated consistently to reflect latest development. Entertainment domain changes rapidly, its webpage has the highest percentage in web page changes; it is likely that the domain name are not retained for a long period of time either however, websites that are science and society related are more static in terms of contents and per page change.

From the reviewed articles, it is found that there are several common elements that comprises a good website. Research criteria has been identified from these elements, even though there are other factors such as design that seems significant as a criteria. It is decided that for the purpose of this research criteria should be quantitatively measurable. Therefore the criteria of this research are divided into two main categories: descriptive and technical. Descriptive criteria contains five factors; domain verification, currency of site, authorship, keyword and description. Technical criteria has website size and download time as determining factor. Domain verification is required to investigate the path to the website address, links that are provided might not direct users to the actual page or the page is not published by the right entity. Domain verification test aims to measure the amount of direct and re-directed sites; direct site has higher credibility since it is registered and has higher accountability to the information posted, re-directed site to a different URL does not conclude that information on that particular site is unacceptable, it merely depreciates slightly the credibility of the site from the user's perception. Re-directed sites affect the speed to access the particular website (Bilal, 2001).

Currency measures the last updated date of the website, websites do not have to be updated frequently but the information must be up-to-date. Currency of a website will depend on the type of the entity it represents and the purpose of the site itself. Timeliness of the information is the currency of the information, the importance of currency varies from types of information it represents; information that represents data has higher priority of currency compared to literature reviews. Since the study is focusing on MSC status companies, it is expected that these companies have well maintained sites and company of this status is expected to have recurring development

information that might have some public or personal interest. A Website that is current shows that it is well maintained and the entity it represents is committed to perform in the industry. With current technology and software development, it is easy for anyone who has minimal computer literacy to create a website, therefore it is important that websites have information on authorship (Mager, 2012).

Although not a critical factor compared to other elements, description and keyword affects the presence of the website on the Internet. Keyword allows search engine to identify the main content of that website which is represented by emphasized words. Internet search engines such as Yahoo! and Google will be able to pick out the sites that contains the word which the user set as search criteria; other unrelated sites that do not contain the keyword will not be included in the search results. The importance of keyword is significant during preliminary query process, search engine will not be able to identify sites that do not contain keywords; in addition to having keywords, it is also important to have the right keyword in order to be identified correctly. After a website has been identified from its keyword, the description of the site will be included in the search result report. This information lets the user know briefly the content of the website before accessing it, for the users, time and unnecessary page visits can be avoided.

In order to evaluate websites further, technical factors such as total size and download time to access is measured. These elements are related in the effect of website efficiency. Size of a site is independent as it depends individually on the developers decision on the objects and elements used. Download time however depends and positively relates to the total size of the website (Xiang & Pan, 2011).

Total size is the amount of data that is needed to be downloaded by the user in order to fully view the web page. Size of a website varies depending on the content and types of object that is on the site, multimedia content such as audio and video is very appealing and increases browsing experience but it is usually large in size. Graphics that is high resolution naturally requires bigger space to be stored in the server. The size of a web page can be reduced with smaller items; menu should be on the client side and avoid tables with graphics as it is slower to load. Rollover GIF files may attract attention but does not contribute to the purpose of the site. Reduce rollover files will reduce the size of the site as well.

Complexity of items affects the size of the site as well, animated GIF, Java script, Java applet and large tables contribute to the complexity of the site. Multimedia enhancement is found to increase popularity of a website. Graphics, pictures, animation and Java applets are common elements that is used to create interactivity, thus increase appeal of users towards the website. This explains the reason entertainment and leisure has the highest graphic quality and content consequently suffers from slow access time due to the size of the data on the site (Hassan, Jones, & Klinkner, 2010).

Even though total size is positively related to the access speed of a website, it is not the only factor. Download time to access can also be affected by the number of connections, server access and data encryption; number of connections and server access are not size related. Limited connection will result in users waiting for available connection once connection ports are exhausted, if objects of a website are located in several different servers, that connection has to be broken and reestablished forever

server connection, during this process download time is increased. In order to secure sensitive information, data is encrypted for transmission, encrypted data relatively larger and does not transmit well on dial-up connection. With the growth of online transactions, a survey conducted by Bell and Tang 1998, found that 30% of retailers who have a website have online transaction facilities and 7% of them charge users for using their transaction facilities. In order to protect this sensitive information that travels in great volume, data must be encrypted. However, for this research, there is no encrypted data used for testing on any of the sample websites. Testing for online transactions was not executed for this study even though samples may have online transaction capabilities (Kammenhuber, Luxenburger, Feldmann, & Weikum, 2006).

It is a standard method how these objects can be merged. It is also a standard for interface, accessing and manipulation. Developers may apply DOM as an interface to their data structures properties and APIs, authors of websites can write DOM standard scripting, not language-specific document thus increases interoperability of the published resource. The main objective of DOM development is to define a programmatic interface for HTML and XML. DOM specification has two parts: Core and HTML, the Core provides low-level set of basic interface that can represent any documents and define additional interface for XML document presentation. However, these interfaces must be implemented with DOM which is the basic of Core section, in contrast to HTML, it does not need to be DOM implemented. DOM implementation provides better convenience in viewing HTML documents in addition to high-level interface between HTML and XML documents.

W3C markup languages focus on the usage of mathematical function in the websites. This is explained in W3C Mathematical Markup Language (MathML™) 2.0 Specification recommendation. This recommendation defines specifically the Mathematical Markup Language, or MathML, which is an XML application that describes mathematical notation and manipulating structure of the document content. The objective of MathML is to enable mathematics to be presented, received and manipulated on the Web similar to HTML text function. The need of this standard arises as two major problems identified for mathematical function usage with HTML document (Jones & Purves, 2008).

The first problem is display, when an equation such as $y = 3.420$ is displayed, the alignment of the horizontal alignment is above the top of the words surrounding the equation. MathML solves this display problem by aligning the equation baseline with the horizontal line of the surrounding words. Figure 2.4 illustrates how MathML is used in a HTML document. The keyword `<mfarc>`, `<msqrt>` and `&PlusMinus` is predefined in the MathML dictionary, therefore there is no coding needed to generate these symbols. MathML may not decrease the size of the sheet with pre-declared keywords, but it manages to display the equation well in the HTML document which is the objective of its development.

The second problem is encoding. It is difficult to search for documents that contains mathematical expression such as $y = 3.420$ and manipulate its mathematical function in a different document. The alt keyboard function may work to copy, but it only captures the equation in a text form, discarding the mathematical function when transferred to a different document. In order to carry on with the research, the primary keyword of this study needs to be

defined. Insights gained from article collection have contributed to the definition of efficiency which is defined as: Accessing information from the website resulted in queries from the right keyword extracted from description in an acceptable time period with author's information. This definition contains the essence of the entire criterion which was mentioned in related research and articles. A website is defined as: A collection of web document that is located in a server that includes a beginning file called the homepage. This document is located in the World Wide Web and represents specific organizations, individuals or interest groups. A website is identified by its address, known as the Unique Resource Locator (URL) which is similar to the hostname (Jones & Purves, 2008).

METHODS

This chapter elaborates the research methodology to investigate the efficiency of MSC status company's website efficiency bases on predetermined criteria. First mentioned in this chapter are research tools; applications and other related peripherals. Next mentioned are the research procedure elaboration and data sub-categorization explanation. Statistical analysis is applied to study the distribution of download speed and the relationship between the sizes of websites with download speed by applying the simple regression method. [WebsiteOptimization.com](http://www.WebsiteOptimization.com) provides free website speed tests without having to install any application on the testing terminal. Websites are analyzed using scripts that are recommended by the Human-Computer Interaction (HCI) Resources. These scripts calculate page size, composition, download time and individual web page components. A URL is required to be keyed into the text field provided and the analysis process will be executed once the Submit Query button has been clicked. A report page will be generated summarizing the analyzed data; total page size in bytes, download times in connection speed of 14.4k, 28.8k, 33.6k, 56k, ISDN 128k and T1 1.44Mbps and recommendation. For this research, data that is collected from the site are: total page size and download time in various connection speeds; the site recommended that pages should be able to be accessed in eight seconds on a 56kbps connection. Application is available at www.websteoptimization.com.

The aim of this study is to determine if the MSC status company's website is effective by measuring the availability of descriptive and technical criteria of the website which are: currency, authorship, keyword, description, loading time and size of the website. First, a search for related journal article, new release and related subject study is conducted. As the articles are identified, the selected publication will be included into the article collection for review. In reviewing literature collection, distinct elements which are highlighted in the article are noted for future reference. Based on distinct elements noted in the literature review, research criteria are determined which is later categorized into two major categories. The two major categories were created based on the nature of the information that is to be acquired. The two major categories are: Descriptive factor and Technical factor (Zhou, Li, & Tang, 2004).

In order to collect data effectively, the website testing application is downloaded and installed accordingly. The application that is required for downloading are Paessler Site Inspector and VIGOS AG Website Analyzer. All the applications were run simultaneously but website testing was conducted consecutively to optimize connection

speed. Collected data were presented in a table with figures and elaboration. Figures and the table were presented according to research major category then followed by category by sectors. Data for currency is further categorized to 30-day period interval; this is done to measure in detail the currency of sites in days. This sub-categorization is also applied to site size and loading time. Site size is categorized to 100kb interval and time to load interval is 20 seconds. The sub categorization is required for data collected from currency, site size and download time to investigate deeper the factor that is affecting the website accordingly. To analyze the efficiency in technical factor, the website size and download time relationship is studied, in order to do this, statistical analysis method of simple linear regression is applied. Simple linear regression analysis is conducted to measure the download speed mean and the relationship with site size. The result from the analysis is used as discussion and conclusion basis in chapter 5 (Tawileh et al., 2010). Overview of the research procedure is summarized below:

1. Literature collection and review
2. Determine research criteria based on literature collection
3. Acquire website testing application to be used as research tools
4. Primary data collection by testing website using research tools
5. Present findings in tables and charts with elaborate description
6. Data from technical category is analyzed using SPSS Analyze-it for Microsoft Excel; add on application to Microsoft Excel, download and install as required.
7. The relationship between size of the website and loading time is determined from the SPSS results.

Discussion and conclusion is derived from the analysis. In order to evaluate the websites further the study of second category which aims to measure technical factors for total size and download time to access is conducted. These elements are related in the effect of website efficiency, size of website is independent as it depends individually on the developers desecration on the objects and elements used. Download time however depends and positively related to the total size of the website. Data type that is collected for both criteria are:

1. Website size – Size of file that is required to view site: Kilobytes (Kb)
Download time – Period of waiting: Seconds(s).

ANALYSIS

The MSC Status Company's website was selected from MSC's web page, an average of thirteen companies were selected from each sector. It is acknowledge that sample of one hundred companies represents a small portion of 1153 MSC status company in the industry. The data was analyzed using SPSS Analyze-It for Excel. Research findings are presented in the form of chart and table followed by data analysis and statistical test mentioned in the previous chapter (Markey, 2007).

From the general observation, data is classified according to sectors; Table 1 presents observation result between sectors. For all sectors, it shows that most websites contains currency information. Consultancy has the highest percentage of last update information on its website at 87%, followed by computers 69%, three sectors shares common percentage of 67% which are communication, software development and Internet business based websites. 64% of hardware development

sector and 62% of life science websites contains currency information.

Table 1: Observation of descriptive factors by sectors

SEE Below Appendix

Authorship factor shows contrasting results, the highest percentage of pages that contains authorship information is 14% from the hardware development sector, double the percentage of communication sector websites. Other sectors such as consultancy, software development, Internet based business, life science and computer's websites do not contain any information for authorship. Overall percentages of description between sectors are relatively low, the highest percentage is from consultancy and communication 27% each, software development and Internet based business is 13%, slightly lower than life science rate which is at 15%. Among the sectors, the computers sector website does not have description for its websites.

The final element for non-descriptive factors is the keyword, consultancy has the highest rate at 33%. Communication is second with the rate at 27%, followed by life science at 15%, hardware development at 14% and at the same rate 13% software development and Internet based business.

Since observation by sector is extracted from general observations, anticipation that data from websites by sector reflects data presented in general observation is realized. Two sectors out of seven have filled all the characteristic of descriptive factor in this study. However the percentage of these elements in sectors are low except for currency where most sector websites do have this information. In addition, it can also be observed that consultancy sector has the highest percentage for all the characteristics that it filled. In general, observation has shown that currency is the highest element that is available in websites of all sectors and authorship is the least.

One of the two elements in the technical factor of this research is loading time of the accessed website. This element corresponds to the arbitrary criterion that websites are to be created with loading speed as a priority. Figure 4.3 presents the general observation of website loading time in seconds, data is classified into 20 second intervals; 0 -20, 21 -40, 41 – 60, 61 -80, 81 – 100, 101 -120 and 120 above with ± 2.878 of error Overview of the observation shows that websites that loads within twenty seconds or less has the highest count at fifty seven and sequentially decreases as loading time increases. Twenty five websites load between twenty one and forty seconds, six websites recorded for loading in the duration of forty one to sixty and sixty one to eighty seconds each (Stenmark, 2008).

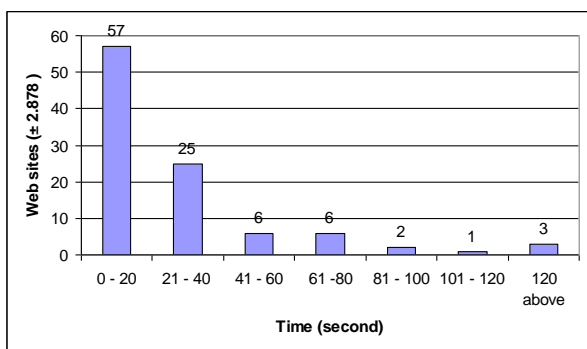


Figure 2: Website loading in seconds

Intervals of eighty one to one hundred, one hundred one to one hundred twenty and one hundred twenty and above, consecutively recorded 2, 1 and 3 websites each, on average from the observation, websites that loads within twenty seconds or less is 8.14 sites per sector. This is a good indication that websites loading in general are relatively fast. To investigate loading time further, general observations are categorized by sector.

In classifying website size according to sector, data from the general observation was divided accordingly. Table 2 shows the distribution of website size according to sector. Life science has 77% of its site size equal or less than ten kilobytes, 15% between 10.1 and twenty kilobytes and only 8% of its site size are between 30.1 to forty kilobytes. This sector does not have site size that is classified in other categories. 54% of websites from the computing sector are sized at ten kilobytes or less, there is no data for the rest of the classification from this sector.

Table 2: Website size by sector.

| Sector/Size | 0 - 10 | 10.1 - 20 | 20.1 - 30 | 30.1 - 40 | 40.1 - 50 | 50above |
|-------------------------|--------|-----------|-----------|-----------|-----------|---------|
| Life science | 77% | 15% | - | 8% | - | - |
| Computers | 54% | 46% | - | - | - | - |
| Consultancy | 60% | 20% | 20% | - | - | - |
| Communication | 40% | 40% | 7% | 7% | - | 7% |
| Hardware development | 43% | 14% | 21% | 14% | - | 7% |
| Software development | 53% | 33% | 7% | 7% | - | - |
| Internet based business | 73% | - | - | 7% | 13% | 7% |

Next sector is consultancy. 60% site sizes are equal or less than ten kilobytes and 20% for both category of 10.1 and twenty kilobytes and 20.1 to thirty. The communication sector has 40% of its websites in the classification of website sizes equal or less than ten kilobytes and 10.1 and twenty kilobytes accordingly. Site percentage that is sized between 20.1 to thirty, 30.1 to forty and fifty kilobytes and above kilobytes is 7%. Similar to life science sector, computer, consultancy, hardware and software development, this sector does not have any data in the four hundred one to five hundred kilobyte category. For hardware development category the percentage distribution of website size are as follows: 43% of its

websites are sized equal or less ten kilobytes, 14% between 10.1 and twenty kilobytes, 21% is 20.1 to thirty kilobytes, 30.1 to forty is 14% and 7% of its website is sized more than fifty kilobytes.

53% of software development website is equal or less than ten kilobytes and 33% is between 10.1 and twenty kilobytes. 20.1 to thirty and 30.1 to forty, both categories has the percentage of 7%. Internet based business has the percentage of 73%, second highest where websites in its sector websites is equal or less than ten kilobytes next to life science. 7% of its website is in the 30.1 to forty and more than fifty kilobytes. The rest of the website which consist of 13% is sized between 40.1 and fifty kilobytes. Authorship is another element of descriptive factor, information for these elements were gathered and projected in Figure 4.6. Figures show the availability of authorship information in the website according to countries, contrasting to previous observation of currency, Malaysia has the least percentage of authorship information available in its websites. United States, UK & Ireland, has equally high percentages where 91% of its web pages contains authorship information. This figure is 51% higher than Malaysia which stands at 40%. Comparatively, Malaysia has the lowest authorship information available for its local websites.

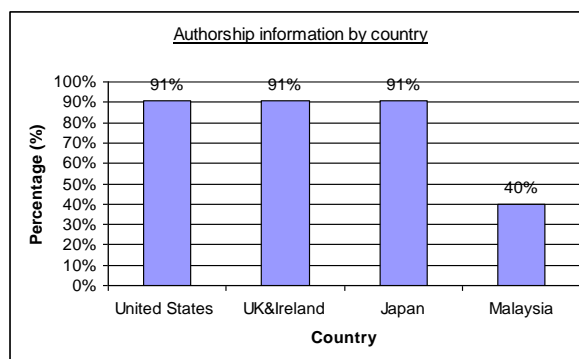


Figure 3: Authorship information by country

The next descriptive factor that is considered is keyword; Figure 4 shows the percentages of websites that contains keyword. United States and Japan has the highest keyword available in its local sites, 73% websites from both countries contains this element. Once again, observation shows that Malaysia has the lowest percentage of element availability, only 22% of Malaysia's website contains keyword. This is less than half of Japan.

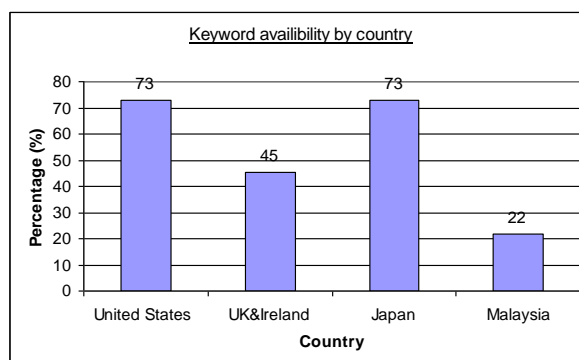


Figure 4: Keyword availability by country

Japan third on the rank has 45% of its websites contains keyword element.

From all the non-technical observation between countries, Malaysia has the least amount of descriptive factor in its website. However, websites from the countries observed Malaysia has the highest percentage of currency availability and United States have the most descriptive elements in its websites. United States has 70.5%, the highest mean of descriptive information that is contain in its website, followed by Japan 68.3%, UK & Ireland 50% and Malaysia 37.25%. In determining this query the relationship between website size and loading speed is investigated. For this query, simple linear regression model is applied; Figure 4.6 shows the model that is used.

$$Y_i = b_0 + b_1X_i \text{ (equation1)}$$

Where

b_0 = Y intercept for the population

Figure 5: Simple linear regression equation

In this model, b_1 , slope of the line represents the expected change in Y per unit change in X. The b_0 which is the Y intercept represents the average value of Y when X is equal to 0. Y_i represents the predicted value as related to changes of X_i where it is the value of X in observation i. To determine the value of b_0 and b_1 the mathematical technique of least-square method is used. The first step in this analysis is to determine the existence of a relationship between site sizes and loading time, this is done by generating a scatter diagram from acquired data. Result from the scatter diagram as presented in Figure 5 shows that data is scattered in a positive linear form. Therefore it can be concluded that there is a positive linear relationship between size of the site and loading time. As size of the site increases, time taken to load also increases. Next is to measure the degree of relationship of the two factors using the simple linear regression model. The mathematical technique of least-square method is used to determine b_0 and b_1 which is achieved by using Analyze-It for Excel. Result summary from the technique is presented in Table 6 below (Chen & Sycara, 1998).

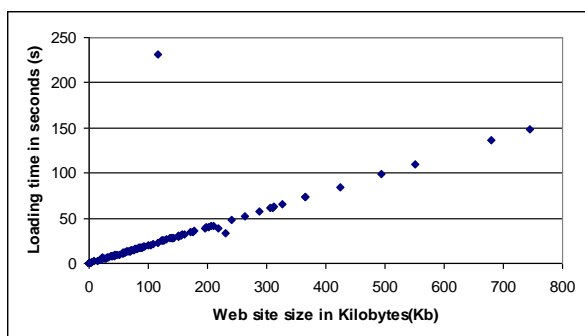


Figure 6: Scatter diagram website size against loading speed

The Y intercept, $b_0 = 2.384$, shows that when data, represented by X, increases by 1kb, the estimated average value of Y is increased by 0.197. This means for each 1kb accessed in size of the website, it is estimated time to increase by 0.197 seconds. Therefore, the slope represents the speed to vary according to site size. The regression coefficient, $r^2 = 0.627$, indicated that 62.7% loading time of website from the sample is dependent to the size of the site. This means that 62.7% of website loading time is exclusively dependent to the size of the site accessed.

DISCUSSION AND CONCLUSIONS

It is a complex problem to evaluate the effectiveness of a web page as there are many elements and approaches to it. Related research in website design shows that websites that incorporated Human-Computer Interaction (HCI) during its development reported higher user satisfaction compared to sites that are purely developed with technical superiority. The study was conducted using questionnaires that was code-named WAMMI (Web Analysis and MeasureMent Inventory) a modified version of SUMI questionnaires which evaluates desktop applications. The WAMMI factors are attractiveness, control, efficiency, helpfulness and learnability. These factors were the main elements in the questionnaire that was used in evaluating the websites. Besides proving the perception that developing website with HCI is more satisfactory to the users, the study was also able to prove that it is possible to measure satisfaction of website in a naturalistic, cost effective environment. It is also possible to compare the perceived usability of different websites in an objective, quantitative manner.

However, this research has undertaken the task of evaluating one hundred and two websites of MSC Status Companies based on authorship, currency, keyword, description, site size and loading speed. Characteristic of research was synthesized from related study and publications. Samples were randomly selected from seven sectors determined by the MSC. Criteria were further grouped into two categories for analysis; descriptive and technical factor. From the analysis, it was found that 69% of websites contain the last updated information, 4% authorship credentials, 18% description information and 22% contains keyword. 57% of websites load within twenty seconds and sized smaller than one hundred kilobytes, the average time of loading is 27.917 seconds on a 56k connection speed.

It is difficult to gauge descriptive factors as it is influenced by independent factors, which most of it is decided by the web developer or the requirement of the organization. The site size is also affected by the same factor, type of object used in the website such as audio, video, graphic and text is determine by the developer of the site. The number of multimedia elements in the website will increase the size of the site but some sites might require high content of these elements. Commercial web pages use multimedia enhancement to increase its website popularity to attract more potential customers. Multimedia enhancement is found to increase popularity of a website graphics, pictures, animation and Java applets are the common elements that is use to create interactivity thus increase appeal of users towards the website. A classic factor that contributes to website popularity is advertisement. It seems that a greater number of advertisement banners increases popularity supporting the notion that advertising banners are successful in promoting a website. Placing advertising or getting listed on a credible Internet

directory will increase popularity and presence of the firm, traditional advertising may have the same effect.

Since it is shown through the simple linear regression method that time taken to access a website is positively dependent on the website size, it is important to maintain a small website. Small size website does not mean that it contains only pure text, graphics can be included but only if it enhances the impact that the information conveys. The core of website development as recommended that the speed, this is to minimize the waiting time to access a website. Waiting on the Internet is defined as devoting more time than perceived necessary in order to complete a task on the Internet Ryan and Valverde (2005). There are fourteen situations where users have to wait for download delay or installing software to continue, these two may occur together. The fourteen situations are generalized into two categories; action and in-action. Action is a situation where users have to wait for something to happen before proceeding and in-action is a situation where the user is interacting with the Internet in order to proceed. The increase usage of broadband will not resolve the issue of waiting on the Internet, as part of the problems are logistic matters such as customer service responds. Since there is lack of effort by firms to measure their website effectiveness, dissatisfied customers have a very high tendency to switch to a competitor or just going to brick-and-mortar stores. Lack of keyword and description are likely to lead to the same result. Potential customers may miss out a particular website if it does not show up on the search engine results since it does not contain description where the keyword can be found.

However, in the case of MSC Status Companies, potential customer may switch to another company for their required product or services. This could cause extensive amount of lost revenue as most companies in the MSC cluster have the resources and potential to serve and provide for global demand. Local market alone is not able to generate enough revenue for the ICT industry. Loss of revenue of the MSC companies indirectly decreases the country's GDP, and if this persists in the long run, the purpose of MSC is no longer relevant. This must be avoided as the government has allocated RM5.2 billion under the eight Malaysia Plan (2001-2005) to incubate and promote the Information Communication Technology (ICT) industry which consists of MSC status companies.

Research Limitation

This study has acknowledge several limitations. First the sample population may not represent the real industry performance. 102 companies as a sample is only 9% out of a total of 1153 MSC status companies that are granted by the MDC. The research criterion that was developed for this study by synthesizing highlighted elements in the literature review as the second limitation. Criterion may not be the accurate factors to be used for measuring a web page. Finally, this study does not reflect the effectiveness of local websites from companies that are not MSC status, in other words it does not represent the overall effectiveness of websites in Malaysia.

CONCLUSION

The second objective of this study is to investigate the efficiency of MSC Status Company's website measured by the technical factors of the web page. Those technical factors are the size of a web page, the time taken to load the web page and comparing the performance of the local websites with websites from the United States, UK & Ireland and Japan. The second objective of this study is concluded as follows:

Loading speed and size of site are tested and determined by the majority of sites which are acceptably fast to load; less than 20 seconds. The result from data analysis shows that there is a positive relationship between the size and loading time of a website. This would mean that a larger website requires a longer time to be accessed. Large websites generally contain more multimedia elements such as animation and sound compared to smaller sites that contain a much simpler form of contents like text and graphics only. Tests to determine the loading time for MSC status company websites less than 20 seconds resulted p-value 0.993 larger than lower critical value which is -1.644, therefore the null hypothesis is accepted. The next test which investigates the website size to be equal or less than 20Kb concluded with the acceptance of the null hypothesis; website from test sample is sized equal or less than 20Kb. The acceptance is due to the p-value which is equal to 1 is bigger than the lower critical value, -1.644. Comparison analysis applying the t-test between websites from United States, UK & Ireland, Japan and Malaysia reveals that there is no difference in mean between any of the country's websites in loading time and size of websites. From this descriptive statistical analysis the entire null hypothesis developed for this test is accepted. This shows that Malaysia's websites have equal performance to foreign websites. The study concluded from all the tests that MSC status companies' websites are efficient.

REFERENCES

1. Bilal, D. (2001). Children's use of the Yahoo!igans! Web search engine: II. Cognitive and physical behaviors on research tasks. *Journal of the American Society for Information science Technology*, 52(2), 118-136.
2. Chau, M., Zeng, D., Chen, H., Huang, M., & Hendriawan, D. (2003). Design and evaluation of a multi-agent collaborative Web mining system. *Decision Support Systems*, 35(1), 167-183.
3. Chen, L., & Sycara, K. (1998). WebMate: A personal agent for browsing and searching. Paper presented at the Proceedings of the second international conference on Autonomous agents.
4. Cheng, A. Y.-n. (2009). Affordances of Learning with Web Search Engines: A Case Study of Undergraduate Students in Hong Kong. *International Journal of Learning*, 16(6), 20-30.
5. Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for web personalization. *ACM Transactions on Internet Technology*, 3(1), 1-27.
6. Hassan, A., Jones, R., & Klinkner, K. L. (2010). Beyond DCG: user behavior as a predictor of a successful search. Paper presented at the Proceedings of the third ACM international conference on Web search and data mining.
7. Hulth, A., & Rydevik, G. (2011). GET WELL: an automated surveillance system for gaining new epidemiological knowledge. *BMC Public Health*, 11(1), 252.
8. Jones, C. B., & Purves, R. S. (2008). Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3), 219-228.
9. Kammenhuber, N., Luxenburger, J., Feldmann, A., & Weikum, G. (2006). Web search clickstreams. Paper presented at the Proceedings of the 6th ACM SIGCOMM conference on Internet measurement.

11. Ma, S., Li, S., & Yang, H. (2016). Utilising Creative Computing and data mining techniques to analyse queries in a meta-search system. Paper presented at the 2016 22nd International Conference on Automation and Computing (ICAC).
12. Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication*
13. *Society*, 15(5), 769-787.
14. Malinský, R., & Jelínek, I. (2010). Improvements of webometrics by using sentiment analysis for better accessibility of the web. Paper presented at the International Conference on Web Engineering.
15. Markey, K. (2007). Twenty-five years of end-user searching, Part 1: Research findings. *Journal of the American Society for Information Science*
16. *Technology*, 58(8), 1071-1081.
17. Stenmark, D. (2008). Identifying clusters of user behavior in intranet search engine log files. *Journal of the American Society for Information Science*
18. *Technology*, 59(14), 2232-2243.
19. Tawileh, W., Mandl, T., Griesbaum, J., Atzmueller, M., Benz, D., Hotho, A., & Stumme, G. (2010). Evaluation of five web search engines in Arabic language. Paper presented at the LWA.
20. Xiang, Z., & Pan, B. (2011). Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations. *Tourism Management*, 32(1), 88-97.
21. Zhou, J., Li, K., & Tang, L. (2004). Towards a fully distributed p2p web search engine. Paper presented at the Proceedings. 10th IEEE International Workshop on Future Trends of Distributed Computing Systems, 2004. FTDCS 2004.
22. De Silva A.D.A., Khatibi A., Azam S.M.F. (2018a). Can parental involvement mitigate swing away from science? Sri Lankan perspectives, *Cogent Education*
23. De Silva A.D.A., Khatibi A., Azam, S. M. F. (2018b). Do the Demographic Differences Manifest in Motivation to Learn Science and Impact on Science Performance? Evidence from Sri Lanka, *International Journal of Science and Mathematics Education*
24. Delafrooz N., Paim L.H., Khatibi A. (2009). Developing an instrument for measurement of attitude toward online shopping, *European Journal of Social Sciences*
25. Dewi N.F., Azam, S. M. F., Yusoff S.K.M. (2019). Factors influencing the information quality of local government financial statement and financial accountability, *Management Science Letters*
26. Doa N.H., Tham J., Khatibi A.A., Azam S.M.F. (2019). An empirical analysis of Cambodian behavior intention towards mobile payment. *Management Science Letters*
27. Maghfuriyah A., Azam, S. M. F., Shukri S. (2019). Market structure and Islamic banking performance in Indonesia: An error correction model, *Management Science Letters*
28. Nguyen H.N., Tham J., Khatibi A., Azam S.M.F. (2019). Enhancing the capacity of tax authorities and its impact on transfer pricing activities of FDI enterprises in Ha Noi, Ho Chi Minh, Dong Nai, and Binh Duong province of Vietnam, *Management Science Letters*
29. Nikhashemi S.R., Paim L., Haque A., Khatibi A., Tarofder A. K. (2013). Internet technology, Crm and customer loyalty: Customer retention and satisfaction perspective, *Middle East Journal of Scientific Research*
30. Nikhashemi S.R., Valaei N., Tarofder A. K. (2017). Does Brand Personality and Perceived Product Quality Play a Major Role in Mobile Phone Consumers' Switching Behaviour? *Global Business Review*
31. Pambreni Y., Khatibi A., Azam, S. M. F., Tham J. (2019). The influence of total quality management toward organization performance, *Management Science Letters*
32. Pathiratne S.U., Khatibi A., Md Johar M.G. (2018). CSFs for Six Sigma in service and manufacturing companies: an insight on literature, *International Journal of Lean Six Sigma*
33. Rachmawati D., Shukri S., Azam, S. M. F., Khatibi A. (2019). Factors influencing customers' purchase decision of residential property in Selangor, Malaysia, *Management Science Letters*
34. Seneviratne K., Hamid J.A., Khatibi A., Azam F., Sudasinghe S. (2019). Multi-faceted professional development designs for science teachers' self-efficacy for inquiry-based teaching: A critical review, *Universal Journal of Educational Research*
35. Sudari S.A., Tarofder A.K., Khatibi A., Tham J. (2019). Measuring the critical effect of marketing mix on customer loyalty through customer satisfaction in food and beverage products, *Management Science Letters*
36. Tarofder A.K., Azam S.M.F., Jalal A. N. (2017). Operational or strategic benefits: Empirical investigation of internet adoption in supply chain management, *Management Research Review*
37. Tarofder A.K., Haque A., Hashim N., Azam, S. M. F., Sherief S. R. (2019). Impact of ecological factors on nationwide supply chain performance, *Ekoloji*
38. Tarofder A.K., Jawabri A., Haque A., Azam S.M.F., Sherief S.R. (2019). Competitive advantages through it-enabled supply chain management (SCM) context, *Polish Journal of Management Studies*
39. Tarofder A.K., Nikhashemi S.R., Azam S. M. F., Selvantharan P., Haque A. (2016). The mediating influence of service failure explanation on customer repurchase intention through customers' satisfaction, *International Journal of Quality and Service Sciences*
40. Udriyah, Tham J., Azam, S. M. F. (2019). The effects of market orientation and innovation on competitive advantage and business performance of textile SMEs, *Management Science Letters*
41. Ulfah R., Amril Jaharadak A., Khatibi A.A. (2019). Motivational factors influencing MSU accounting students to become a certified public accountant (CPA), *Management Science Letters*

Appendix

| Table 1: Observation of descriptive factors by sectors | | | | |
|---|----------|------------|-------------|---------|
| | Currency | Authorship | Description | Keyword |
| Consultancy | 87% | - | 27% | 33% |
| Communication | 67% | 7% | 27% | 27% |
| Hardware development | 64% | 14% | 7% | 14% |
| Software development | 67% | - | 13% | 13% |
| Internet based business | 67% | - | 13% | 13% |
| Life science | 62% | - | 15% | 15% |
| Computers | 69% | - | - | - |