

Evaluating the Usability and Content Usefulness of Web Sites: A Benchmarking Approach

Shahizan Hassan, Northern University of Malaysia, Malaysia
Feng Li, University of Newcastle upon Tyne Business School, UK

EXECUTIVE SUMMARY

Although benchmarking technique has been widely used in various aspects of organisations and businesses, there is no clear framework on how the technique can be applied for Web evaluation. This article presents a framework for evaluating the usability and content usefulness of Web sites by using the benchmarking approach. It describes the purpose of evaluation, metrics to be used, and processes through which Web benchmarking can be carried out. Several methods were used which include content analysis of literature and expert review. A total of 46 criteria were identified that can be used as the benchmarking metrics. The framework was tested for its applicability by evaluating four political Web sites in Malaysia. The results proved that the framework is easy to implement and would be particularly valuable for those who intend to benchmark the overall usability and content usefulness of their Web sites against those of their competitors.

Keywords: *benchmarking; evaluation; usability; Web content; Web design*

INTRODUCTION

Benchmarking is a measuring method widely used by companies to improve many areas of activities including human resource management, information systems, customer processes, quality management, purchasing, and supplier management (Elmuti, 1998). The common goal of this approach is to identify the 'best practices' of other organisations so that it can be implemented in one's own operation. In Web evaluation,

benchmarking could be used to measure the performance of one's Web site against others, especially its competitors. By doing this, strengths and weaknesses of one's Web site can be identified, and the quality and usefulness of the Web site could be improved accordingly. For many years, the benchmarking technique has proven its success and widely been used in business (Government Centre for Information Systems, 1995) and various aspects of organisations. However, *very little information* is avail-

able on how this approach can successfully be implemented in Web site evaluation.

With this in mind, a framework was developed on how the benchmarking technique can be applied to measuring Web sites in terms of their usability and content usefulness. This framework is aimed at both technical and non-technical people who are involved in Web site design and evaluation. Some empirical work was conducted to test the applicability of the framework which will be presented in this article.

The article will first describe some existing Web evaluation methods, followed by the definition of the concept of Web usability and content usefulness. Then methods used in this study will be explained briefly. Next, the findings and the proposed benchmarking framework are then discussed in detail. Finally, the article ends with some suggestions for future studies.

EXISTING WEB SITE EVALUATION METHODS

Despite the lack of Web evaluation studies that use the benchmarking technique, many studies on Web evaluation have been carried out for many years which employ conventional methods, including usability testing (e.g., Nielsen, 1993; Zimmerman, 1998), expert review (e.g., Shneiderman, 1998; Zhang & Dran, 2000), case study (e.g., Smith, Newman, & Parks, 1997), and automated assessment (e.g., Tuasher & Greenberg, 1997; NetMechanic, 2000).

However, several attempts were made to measure Web sites using the benchmarking approach. Simeon (1999), for example, performed a study on how benchmarking techniques can be used to compare the Attracting, Informing, Posi-

tioning, and Delivering (AIPD) strategies of commercial Web sites in order to clarify strategic opportunities and advantages. In this study, he used the AIPD approach to compare Web site strategies of 68 American and 54 Japanese banks. Nonetheless, this approach has its limitations in that there is no clear explanation on how the AIPD elements were identified and grouped into the four (AIPD) categories. In addition, the AIPD model is only applicable to banking Web sites, and no attempts have been made yet to test it on other types of Web sites. Another study was carried out by Misic and Johnson (1999), where four factors of Web site effectiveness (functions, navigation, content, and contact information) were used to benchmark the Web site of the College of Business (COB) at Northern Illinois University (NIU) against 45 other business schools. The main limitation of this study is the lack of items used in the metrics. It only covers limited aspects of functional/navigational issues, content and style, and contact information. Other important aspects of Web evaluation, such as proper use of multimedia elements and issues of accessibility, are not included.

There are also attempts to design methodologies for benchmarking and assessing the quality of Web sites. A good example is WebQual, a method for measuring the quality of an organisation's e-commerce site (Barnes & Vidgen, 2002). WebQual uses an index that gives an overall rating of a Web site based solely on customer perceptions of quality weighted by importance. The index was designed according to three dimensions — usability, information quality, and service interaction quality. Although the measures were claimed to be valid and reliable, the main limitation of WebQual is that the results heavily rely on the Internet users' percep-

tions. Other actors — for example, designers, business partners, suppliers, and related governmental organisations — are not considered. In addition, the method has so far been applied only to business firms, for example, the Internet bookshops, thus raising the question of whether the criteria used can also be applied to non-commercial Web sites. WebTango is another Web evaluation methodology worth mentioning. It is an automated tool developed specifically for assessing Web usability and accessibility (Ivory-Nduaye, 2003). This tool is very useful in the sense that it can be used by designers to measure their Web sites automatically. However, the tool focuses more on design aspects rather than content.

From this, it is clear that although the benchmarking approach has been successfully used in many areas of business functions, the applicability of this approach in Web site usability and content evaluation requires further research.

CONCEPT OF USABILITY

Usability is a very broad concept in system design and is defined differently by different HCI scholars. Because of this, approaches to measuring usability also differ between each other. Shakel (1991), Nielsen (1993), and Lu and Yeung (1998) defined usability as an attribute to a product or system acceptance. Therefore, their model of usability is explained in terms of its relationship with the concept of acceptability of a system.

According to Shakel, a user would compare four properties of a system to the sacrifices needed to use it (Shakel, 1991). The properties are *utility* (the match between users' needs and functions of a product), *usability* (users' ability to utilise the functions), *likeability* (users' affective

evaluations), and *costs* (both financial and social consequences). In view of this, Shakel defines a product or system's acceptance as a function of perceived utility, usability, likeability, and costs. Nielsen (1993) on the other hand, presents a slightly different concept by suggesting that both usability and utility together form usefulness. Utility refers to whether the functionality of a system can do what is needed, while usability relates to the question of how well users can use the functionality. Usefulness and other perceived system attributes like cost, compatibility, and reliability will lead to practical acceptability of a system. In defining usability, Nielsen outlines five operational criteria: *learnability* (the novices' ability to reach a reasonable level of performance), *efficiency* (expert users' level of performance), *errors* (users' number of errors), *satisfaction* (users' subjective assessment of a system), and *memorability* (users' ability to remember how to use a system). Nielsen's view is supported by Lu and Yeung (1998), who also propose the concept of usefulness as one of the attributes for system acceptability. Their model divides usefulness into two aspects — functionality (which is similar to utility in Nielsen's model) and usability. Lu and Yeung's perception of usability is slightly different since their model is developed specifically for the Web environment. Usability in this model refers to users' ease-of-browsing, ease-of-reading, and satisfaction.

The usability models described here highlight the need for researchers to understand the underlying concept of usability as a measurement. Although the approaches towards defining the concept of usability are slightly different between each other, all models tend to have agreement on the dimensions of usability that cover

aspects of effectiveness, efficiency, learnability, and user satisfaction. One important issue with regard to usability definition is the question of whether content coverage of a system should be included as one of the elements of usability. To date, there is no clear discussion on this in the usability literature. However, most models of usability (Shakel, 1991; Nielsen, 1993; Lu & Yeung, 1998) include 'user satisfaction' as one of the usability criteria. This element has an indirect relationship with the need for content quality of a particular system. User satisfaction is related to users' subjective assessment on a particular system in terms of its ease of use, as well as its usefulness. Thus, it can be said that both user interface and content together determine users' level of satisfaction.

METHODOLOGY

There are five phases of this study: (1) the identification of the metrics for benchmarking Web usability and content usefulness from the literature, (2) the verification of the metrics from the usability experts, (3) the classification of the objective and subjective criteria, (4) the development of the benchmarking framework, and (5) the application of the framework on selected political Web sites in Malaysia as an example to test the framework.

Content analysis used in phase 1 was to analyse various literature on Web usability. The main objective was to gather the key Web usability criteria proposed in the selected literature. These criteria will then be used as the metrics for benchmarking the usability of Web sites. Several guides, articles, and textbooks were selected based on the recommendation from the human-computer interaction (HCI) scholars such as Jakob Nielsen

(useit.com, 2002), Keith Instone (usableWeb.com, 2002) and Gary Perlman (Perlman, 2001). At least 30 guides were selected including *IBM Web Design Guide* (IBM, 2000), *Yale Style Manual* (Lynch & Horton, 1999), *Microsoft Web Workshop* (Web Workshop, 1999), *Designing Information-Abundant Web Sites: Issues and Recommendations* (Shneiderman, 1997), *Writing for the Web Guide* (Sun Microsystems, 1999), *Web Content Accessibility Guidelines 1.0* (W3C, 2002), *Web Graphics Design* (Benjamin, 1996), and *Web Design: The Complete Reference* (Powell, 2000).

All these guides were analysed to extract generic criteria of Web usability. Each criterion identified was recorded in a standard form. During the analysis, all criteria and elements of usability that were considered too technical were rephrased, or in some cases, excluded to cater for both technical and non-technical people. Both concept existence and frequency were used for coding the data (i.e., any usability criterion identified from the selected literature was not only coded for its existence, but also for the frequency it was mentioned — see example in Table 1). Once all the criteria were recorded, they were revised and refined to remove duplication. Then, the criteria were grouped into seven categories based on suitability and context of use, which will be discussed further in the next section. The outcome from phase 1 was a revised list of Web usability criteria grouped into seven categories.

During the second phase, the expert review method was utilised to verify the key criteria identified from the literature in phase 1. A list of the identified criteria was sent to 36 experts for review and verification; 15 replied with their comments and review. These experts are those who have

Table 1. An example of data analysis summary for Web criteria elicitation

No.	Criteria	Guide 1 (Benjamin, 1996)	Guide 2 (IBM, 2000)	Guide 3 (Web Workshop, 1999)	Guide 4 (MIT et al., 1999)	Total
1	Clear title for each page	X	X	X	X	4
2	Compatible content for all main browsers	X	X	-	-	2
3	Labelling of all static media	-	X	X	-	2

Note: 'X' means that the criterion was mentioned in the guide.

more than five years of experience in HCI and usability areas. The selection of experts involved two processes — identifying the experts from the proceedings of past conferences on Human Factors in Computing Systems (CHI) and sending invitations to participate in the CHIWeB e-mail list. The main objective of the review was to get verifications and suggestions from the experts with regard to generic Web usability criteria. Additionally, they were also requested to comment on the suitability of the criteria groupings. The experts were allowed to edit (add, delete, and rephrase) all the criteria derived from the literature in a review form (see example in Table 2).

Their feedback was analysed and used to further refine the usability list.

During phase 3, all criteria were analysed and classified into objective or subjective measures. A two-hour brainstorming session involving three evaluators was carried out for this purpose. Since Web design environments are closely related to multimedia, information retrieval, and networking areas (Powell, 2000), the three selected evaluators were those who had strong knowledge in each of these disciplines. Card sorting technique was used during the brainstorming session to classify the criteria. Using the criteria derived from phase 3, a benchmarking framework

Table 2. An example of the expert review form

Criteria for Navigation	Reviewer's Comment	
Objective	About criteria	About Grouping
1. Menus are fit on screen (no scrolling)	OK	OK
2. Accurate/unbroken links	OK	OK
3. Menus are positioned on the left side of screen	Not a recommended criterion. Would be a disadvantage for left-handed users.	-
4. Use of text within text link (where applicable)	OK	OK
5. Labelling of all static media used as menu	OK	Should be moved to 'Media Use' group

was developed in phase 4. Several general models for benchmarking were referred to, including Chang and Kelly (1995), Codling (1992), Bramham (1997), and Anderson (1996).

Finally, in phase 5, the framework was tested on all major political Web sites in Malaysia for its applicability and practicality. Political Web sites were selected for this study as they are non-commercial Web sites that deal with the government and the general public (customers). Studies on commercial Web sites are plenty (business-to-customers), but only a few focus on non-commercial sites, particularly political Web sites (government-to-consumers). Furthermore, at the time of testing in phase 5 (late 2002), Malaysia was about to have a general election. During this time, most political parties heavily utilise Web sites as one of their political communication media.

WEB USABILITY AND CONTENT USEFULNESS: MAIN FINDINGS

Content analysis of the literature and expert review had resulted in the identification of 57 key criteria of Web usability, which were clustered into seven main

groups based on their suitability and context as shown in Table 3.

Screen Appearance

Screen appearance or layout can be divided into four categories — space provision, choice of colour, readability, and scannability (Lynch & Horton, 1999; Seminerio, 1998). All experts agree that these are four very important areas of usability. More space should be allocated for contents, and the variety of different screen types (palm tops, television, etc.) should be taken into consideration. Additionally, proper use of colour not only attracts users to visit a Web site, but also improves learnability and ease of use. Equally important is the issue of readability. Almost all experts agree that a readable content is associated with choice of colour, fonts, and the use of colour for text and background. Apart from that, designers should not only design for readability, but also for scannability, for example, the use of typography and skimming layout. The proposed list of Web usability criteria for Screen Appearance is presented in Table 4.

Table 3. Seven categories of Web usability criteria

<i>Grouping of Web Criteria</i>	<i>No. of Criteria</i>		
	<i>Objective</i>	<i>Subjective</i>	<i>Total</i>
Screen Appearance	8	2	10
Content	16	2	18
Accessibility	3	1	4
Navigation	8	3	11
Media Use	5	3	8
Interactivity	3	0	3
Consistency	3	0	3
Total	46	11	57

Table 4. List of Web usability criteria for screen appearance

<i>Sub-Category</i>	<i>Objective</i>	<i>Subjective</i>
Space allocation	-NA-	<ul style="list-style-type: none"> More space for contents than to other display elements
Choice of colour	<ul style="list-style-type: none"> Non-excessive use of colour for text Sharp colour contrast between text and its background Use of colour to differentiate between functional area (e.g., tool bar and menu bar) with content display 	-NA-
Readability	<ul style="list-style-type: none"> Different text sizes to differentiate between titles, headings, and texts Avoidance of background images in the content display area 	<ul style="list-style-type: none"> Use of fonts that are easy to read
Scannability	<ul style="list-style-type: none"> Clear titles for each page Clear headings, sub headings for text/document Use of typography and skimming layout (e.g., bold fonts and highlighted words) 	-NA-

Accessibility

One of the goals of having a Web site is to attract as many visitors as possible from various locations. The basic way to achieve this is to ensure that the site is accessible to the target users. In this study, three elements of accessibility are identified — *loading time*, *browser compatibility*, and *search facility*. Generally, users could not tolerate long loading time (Morkes & Nielsen, 1999). As such, designing for speed should be a priority for the designers. The *Yale Style Manual* (Lynch & Horton, 1999) ranks “design for speed” as top priority by stating that the threshold of frustration for most computing tasks is around 10 seconds. All experts agree that users should not be kept too long while waiting for a Web page to load. However, they failed to have any agreement on the exact length of waiting time that would be considered acceptable by the users. Apart from loading time, designers should also consider different browsers with dif-

ferent versions used by the Internet users across the world. Additionally, the experts also agree on the need to provide an effective local search facility, because it will speed up users’ searches for information on a particular Web site. One of Nielsen’s studies found that search facility is highly recommended by the participants (Nielsen, 1997a). The proposed accessibility elements of the Web are shown in Table 5.

Navigation

Some people believe that the best site should contain a lot of graphics, animation, and colour, but often neglect a basic element of an effective Web site: its navigability. Good navigation in a Web site is comparable to a good road map. Our findings of the expert review show that with good navigation such as logical tree-like structure, proper grouping of contents, and use of navigational tools on all pages, users know where they are, where they have been, and where they can go from their

Table 5. List of Web usability criteria for accessibility

<i>Sub-Category</i>	<i>Objective</i>	<i>Subjective</i>
Loading speed	-NA-	• Acceptable loading time
Browser compatibility	<ul style="list-style-type: none"> • Compatible contents for all main browsers • Compatible contents between different versions of the same browser • Compatible display for different screen types (e.g., black & white, palm top, and digital TV) 	-NA-
Search facility	<ul style="list-style-type: none"> • The use of local search facility 	-NA-

Table 6. List of Web usability criteria for navigation

<i>Objective</i>	<i>Subjective</i>
<ul style="list-style-type: none"> • Menu/list of key content in the main page • Menu/list of key content in all sub-pages* • Links to the main page in all sub-pages* • Accurate/unbroken links • Use of sitemap • Menus are fit on screen (no scrolling) • Use of text within text link (where applicable) • No/short page scrolling 	<ul style="list-style-type: none"> • The wording for each category of contents is meaningful to users • Contents should be grouped into a small number of key categories • Small number of steps/links to arrive at a particular information (rule of thumb is 3)

* These criteria are not applicable to Web sites that use frames/separate windows for sub-pages.

current position. In short, navigation is the key to making the experience enjoyable and efficient. The usability criteria proposed by the experts are listed in Table 6.

Media Use

The main multimedia elements are sound, graphics, images, audio, and video (Shirley, 1999). Some Web sites embed audio as background music, downloadable audio files, or on-the-fly audio clips. Sound may also be used in conjunction with animation or video. As with colour, sound can help improve or degrade usability. There are things that cannot be described by words, and thus the use of graphics and images is very helpful. Furthermore, in certain cases graphics could be used to

emphasise text. In our study, all experts emphasise the need for providing alternative access of information whenever audio, animation, and video elements are used to allow accessibility for those having browsers that do not support the elements. Table 7 presents the list of the proposed usability criteria for proper use of media.

Interactivity

Interactivity is a broad concept. In this study, it refers to *features in a Web site that facilitate a two-way communication between users and site owners or other pre-assigned personnel*. Additionally, the features allow users to give feedback and comments on issues raised by the Web site. The introduction of the interactivity features

Table 7. List of Web usability criteria for media use

<i>Sub-Category</i>	<i>Objective</i>	<i>Subjective</i>
Continuous/time-based media (audio, animation, and video)	<ul style="list-style-type: none"> • Control features for continuous media where appropriate (e.g., replay/turn off) • Alternative access (e.g., text version) to any information in continuous media • Avoidance of looping animation to prevent users' distraction 	<ul style="list-style-type: none"> • Use of continuous media to suit content (e.g., demonstration, instruction, and speeches)
Static media (graphics, images, pictures)	<ul style="list-style-type: none"> • Labelling of all static media, especially those used for menus and icons • Use of thumbnails to display photos 	<ul style="list-style-type: none"> • Use of static media to enhance the information being presented • Non-excessive use of static media

such as e-mail, guess book, and net forum may enhance a Web site's worthiness. While agreeing that these elements are important, some of the experts say that making them available is insufficient. Designers should take into consideration whether the elements are effective and easy to use, especially when dealing with multiple forms. Three criteria are proposed and agreed by the experts as presented in Table 8.

Consistency

This study also found that design consistency is important to speed up users' learning. All experts agree to the fact that designers need to provide consistent lay-

out for title, subtitle, page footers, background, and navigation links and icons in terms of colour, size, space, and fonts used. However, one of the experts suggests that minor changes should be made to the structure of the screen appearance every now and then so that users will not get bored and banner blind. Details on the proposed usability criteria for consistency are shown in Table 9.

Content

Apart from user interface, content is undoubtedly another very important element of Web sites. It is the content that attracts people to visit a particular Web site. Among the suggested criteria by the ex-

Table 8. List of Web usability criteria for interactivity

<i>Objective</i>	<i>Subjective</i>
<ul style="list-style-type: none"> • Availability of features for users' feedback about the site (e.g., Web master's e-mail address and online form) • Availability of features for sharing views and discussions (e.g., e-forum, net conference, and net chatting) • Availability of entertainment features (e.g., online games and puzzles) 	-NA-

Table 9. List of Web usability criteria for consistency

<i>Objective</i>	<i>Subjective</i>
<ul style="list-style-type: none"> • Consistent page layout (e.g., screen size for content display, banners, and menu bar) • Consistent use of text in terms of its type, font size, and colour • Consistent use of navigational aids (e.g., menu bar, buttons, and links in terms of graphics metaphor, size, and colour) 	-NA-

perts are suitable language for audience, high-quality writing with no grammatical and typographical errors, passages that are easy to read and understand, clear information about authors, and references cited where applicable. In addition, several experts suggest that merely having a section for press release and publication is not enough. Instead, Web developers should ensure that these publications are up to date

and being archived accordingly. There is also a suggestion that users are to be informed about the difference between internal and external links. Providing a printer-friendly environment within Web pages that offer long documents could also boost usability. The result of the expert review pertaining to the generic criteria of content usefulness is shown in Table 10.

Table 10. List of generic criteria for content usefulness

<i>Sub- Category</i>	<i>Objective</i>	<i>Subjective</i>
Scope	<ul style="list-style-type: none"> • Suitable language for audience • Up-to-date publication (e.g., news & articles) • Archive of previously published materials 	<ul style="list-style-type: none"> • Contents provided meet the expectation of target users
Accuracy	-NA-	<ul style="list-style-type: none"> • High-quality writing (e.g., good grammar)
Authority and Reliability	<ul style="list-style-type: none"> • Information on authors of text/documents (e.g., names) • References or sources of text and other documents • Background information of institution/organisation/owner of the site (i.e., logo, name, address, phone number) 	-NA-
Currency	<ul style="list-style-type: none"> • Up-to-date contents 	-NA-
Uniqueness	<ul style="list-style-type: none"> • Options for output/print format when appropriate • Choices of language for multi-ethnic audience • Choices of media type for particular information (e.g., text only, audio, or video) • Information/warnings on file type and size for downloading 	-NA-
Linkages	<ul style="list-style-type: none"> • Clear distinctions between internal and external links • Links to other relevant sites (e.g., state and local branches) 	-NA-
Text Quality	<ul style="list-style-type: none"> ■ News/articles/documents/stories with pictures ■ Summary of news/articles/documents/stories with links to full versions ■ Divide news/articles/documents/stories according to scope (e.g., local and international) 	-NA-

EVALUATION: BENCHMARKING APPROACH

This section explains our framework for benchmarking the usability and content usefulness of Web sites. As presented earlier, the identified evaluation list was grouped into objective and subjective criteria. In this framework, only the objective criteria will be used as the benchmarking metrics because they are absolute criteria that can be measured easily even by laypeople. In contrast, the subjective criteria are mostly relative measures that can only be evaluated qualitatively. Furthermore, these criteria depend on the perception of users towards a particular Web site. The main purpose of this framework is to assist individuals or teams that intend to measure the usability of their Web sites against those of their competitors or of similar types. It provides guidance to technical and non-technical people who are involved in Web evaluation projects on what, who, and how to benchmark Web sites. This framework can also be used by those who want to know the generic usability criteria that need to be taken into account in determining the level of Web usability.

What is Benchmarking?

Benchmarking is comparing ones' current performances and practices with others in the same area of interest or business (Codling, 1992; Bramham, 1997). The result of benchmarking is normally used for bridging the gap with competitors and moving from where one is now to where one wants to be (Chang & Kelly, 1995). There are many advantages an organisation could gain from benchmarking, including creating awareness of changing consumer needs

and enabling improvements through learning from others who are better. Benchmarking can also be performed on Web sites and can be divided into two types (Anderson, 1996; Bendell, Boulter, & Kelly, 1993)—internal benchmarking (comparisons between Web sites of units/departments/branches) and competitive or external benchmarking (direct comparisons against competitors' Web sites outside an organisation).

Eight Steps to Web Benchmarking

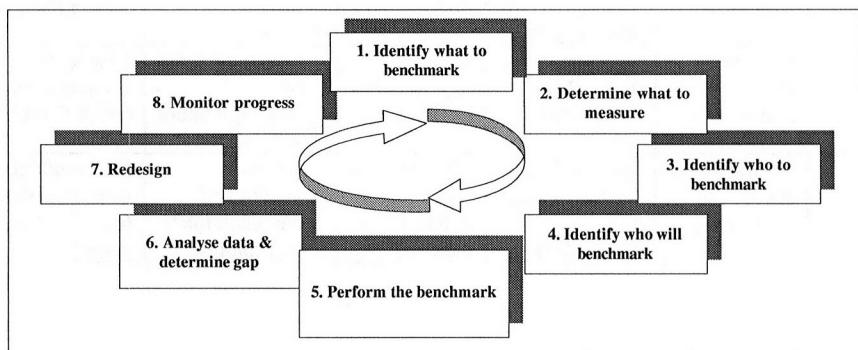
Web benchmarking is a continuous process of measuring and comparing one's Web sites with others, which involves at least eight steps (Chang & Kelly, 1995; Codling, 1992; Bramham, 1997; Anderson, 1996) as shown in Figure 1.

Step 1: Identify What to Benchmark

There are many aspects of Web sites which can be evaluated in order to improve their effectiveness and usefulness. One of them is usability—the main focus of this framework. As explained earlier, Web usability is a broad concept covering at least seven major factors—Screen Appearance, Consistency, Accessibility, Navigation, Media Use, Interactivity, and Content. Considering this, those who intend to benchmark their Web sites need to decide whether to benchmark all seven factors or only concentrate on certain factors. This decision depends on the purpose of the evaluation, time constraint, and the number of people involved.

Step 2: Determine What to Measure

Once factors of Web usability to benchmark have been decided, one needs

Figure 1. Eight steps to Web benchmarking

to determine what measures to use for each factor. Table 3, presented earlier, provides the information on the number of measures (objective criteria) to be used in the benchmarking for all factors. Although Web sites can be measured quantitatively or qualitatively, this framework only focuses on the quantitative measures by using all 46 objective criteria as listed in Table 4 through Table 10.

Step 3. Identify Benchmarking Sites

The identification of Web sites to benchmark depends on the type of benchmark to be performed (i.e., internal or external benchmarking). When performing internal benchmarking, one shall select Web sites of other departments or branches within the same organisation. On the other hand, for an external benchmarking, one shall select a number of Web sites (at least three) of closest competitors. In the case of too many Web sites to benchmark with, one can employ a suitable sampling technique.

Step 4. Select Evaluators

Selecting evaluators will not be difficult because they do not have to be experts in HCI or Web usability areas. The metrics used in the benchmarking are based on Web usability criteria that are easily understood by general Internet users. Evaluators could be any individuals who are competent and frequent Internet users, and who are familiar with the Web environment and terminology. They should be independent (not members of the design team) so that the issue of potential bias can be avoided. The number of evaluators to be used in the benchmarking will depend on the timeframe and budget provided for the Web evaluation project. However, for quick and better results, at least two evaluators should be selected.

Step 5. Perform the Benchmark

Once one has identified what to benchmark (step 1), what measures to use (step 2), the benchmarking sites and the evaluators (steps 3 and 4), the

Table 11. Computer specifications for Web benchmarking

<i>Computer</i>	<i>Processor</i>	<i>Random Access Memory (RAM)</i>	<i>Internet Access</i>	<i>Internet Browser</i>	<i>Screen Resolution</i>
1	Low speed (e.g., Pentium 1)	Low memory (e.g., 32 megabytes)	Low speed (e.g., 33Kbps modem)	Lower versions of Netscape/Internet Explorer	Low screen resolution (e.g., 640 X 480 pixels)
2	Medium/high speed (e.g., Pentium 3)	Medium/high Memory (e.g., 128 megabytes)	Medium/high speed (e.g., 56Kbps modem)	Higher versions of Netscape/Internet Explorer	Medium/higher screen resolution (e.g., 1024 X 768 pixels)

benchmarking process can be conducted. First, prepare the necessary equipment and a suitable room for the benchmarking. A minimum of two computers should be used with the specifications described in Table 11.

Computers with different specifications, network connection capabilities, Internet browser versions, and screen resolutions as described above are necessary to assess different usability aspects of Web sites, particularly those related to display compatibility. Furthermore, not all Internet users are using the latest computer technology with high specification. When the

necessary equipment is ready, a briefing session on the purpose of the benchmarking and how to carry it out should be given to the evaluators. Each evaluator should be provided with a set of benchmarking forms (for each Web site), which contain the list of Web usability criteria as listed in Tables 4 through 10 (only objective criteria). The evaluators will then fill in the form (see a sample form in Table 12) while assessing the selected Web sites. They will tick "YES" for criteria *existence* and "NO" for *non-existence*. For the media use category, an additional column should be provided for

Table 12. Sample of the benchmarking form

Web Usability Benchmarking			
URL:	Date:	Time:	
Evaluator's Name:		Category: Screen Appearance	
Subcategory	Criteria	YES	NO
Choice of colour	1. Non-excessive use of colour for text 2. Sharp colour contrast between text and its background 3. Use of colour to differentiate functional area with content display area	()	()
Readability	4. Different text sizes to differentiate between titles, headings, and texts 5. Avoidance of background images in the content display area	() ()	() ()
Scannability	6. Clear titles for each page 7. Clear headings, sub-headings for text/document 8. Use of typography and skimming layout (e.g., bold fonts and highlighted words)	() () ()	() () ()
Total		(/8)	(/8)

Copyright © 2005, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Table 13. Summary form for the number of usability criteria in all SCANMIC categories

	Web Site A			Web Site B			Web Site C		
	YES	NO	NA	YES	NO	NA	YES	NO	NA
<u>Screen Appearance (F1)</u>									
<u>Choice of colour (/3)</u>									
<u>Readability (/2)</u>									
<u>Scannability (/3)</u>									
<u>Total (/8)</u>									
<u>Consistency (F2) (/3)</u>									
<u>Accessibility (F3)</u>									
<u>Display compatibility (/2)</u>									
<u>Search facility (/1)</u>									
<u>Total (/3)</u>									
<u>Navigation (F4) (/8)</u>									
<u>Media Use (F5)</u>									
<u>Static media (/3)</u>									
<u>Continuous media (/2)</u>									
<u>Total (/5)</u>									
<u>Interactivity (F6) (/3)</u>									
<u>Content (for general Web sites (F7)</u>									
<u>Scope (/3)</u>									
<u>Authority and reliability (/3)</u>									
<u>Currency (/1)</u>									
<u>Uniqueness (/4)</u>									
<u>Linkages (2)</u>									
<u>Text Quality (/3)</u>									
<u>Total (/16)</u>									
Grand Total (/46)									

“NA” (Not Applicable) because not all Web sites fully utilise all media elements.

Step 6. Analyse Data and Determine the Gap

The next step is to analyse the data derived from step 5. First, the data can be summarised by counting the number of *existence* (YES) and *non-existence* (NO) of the criteria for each SCANMIC category. An example of a form that can be used for this is shown in Table 13. Then, the data can be further analysed to identify the usability level of the Web sites that are being benchmarked as exemplified in Table 14.

From the analysis, the gaps that exist between the Web sites can be determined. If $U_B > U_A$ and $U_B > U_C$, then Web site B shows higher usability level than Web sites A and C. To more clearly see the gaps, plotting charts such as bar charts based on the results can be used.

Step 7. Redesign

The results derived from steps 5 and 6 will help identify weaknesses and strengths of one’s Web site against others in terms of usability. In particular, areas of concern that need to be modified and enhanced can be located. Due to the dynamic nature of Web sites and the Internet tech-

Table 14. An example of calculation for percentage Web usability index for 3 general Web sites

SCANMIC Factors	Web Site A			Web Site B			Web Site C		
	YES	NO	NA	YES	NO	NA	YES	NO	NA
F1 (/8)	a _A	h _A		a _B	h _B		a _C	h _C	
F2 (/3)	b _A	i _A		b _B	i _B		b _C	i _C	
F3 (/3)	c _A	j _A		c _B	j _B		c _C	j _C	
F4 (/8)	d _A	k _A		d _B	k _B		d _C	k _C	
F5 (/5)	e _A	l _A	o _A	e _B	l _B	o _B	e _C	l _C	o _C
F6 (/3)	f _A	m _A		f _B	m _B		f _C	m _C	
F7 (/16)	g _A	n _A		g _B	n _B		g _C	n _C	
Total (/46)	$\Sigma a_A + \dots + g_A$	$\Sigma h_A + \dots + n_A$	Σo_A	$\Sigma a_B + \dots + g_B$	$\Sigma h_B + \dots + n_B$	Σo_B	$\Sigma a_C + \dots + g_C$	$\Sigma h_C + \dots + n_C$	Σo_C
	YES _A	NO _A	NA _A	YES _B	NO _B	NA _B	YES _C	NO _C	NA _C

$$\% \text{ Usability Index for Web Site A} = U_A = \left\{ \frac{\text{YES}_A}{46 - \text{NA}_A} \right\} \times 100 \quad \text{or} \quad = U_A = \left\{ \frac{46 - \text{NO}_A}{46 - \text{NA}_A} \right\} \times 100$$

$$\% \text{ Usability Index for Web Site B} = U_B = \left\{ \frac{\text{YES}_B}{46 - \text{NA}_B} \right\} \times 100 \quad \text{or} \quad = U_B = \left\{ \frac{46 - \text{NO}_B}{46 - \text{NA}_B} \right\} \times 100$$

$$\% \text{ Usability Index for Web Site C} = U_C = \left\{ \frac{\text{YES}_C}{46 - \text{NA}_C} \right\} \times 100 \quad \text{or} \quad = U_C = \left\{ \frac{46 - \text{NO}_C}{46 - \text{NA}_C} \right\} \times 100$$

nology, redesigning Web sites has become a continuous process. As such, the result from the benchmarking should also be of help for Web developers in their Web redesign process. In some cases, Web redesign and enhancements would require more money and manpower. Thus, the benchmarking results can also be used to justify the need for more funding in Web projects.

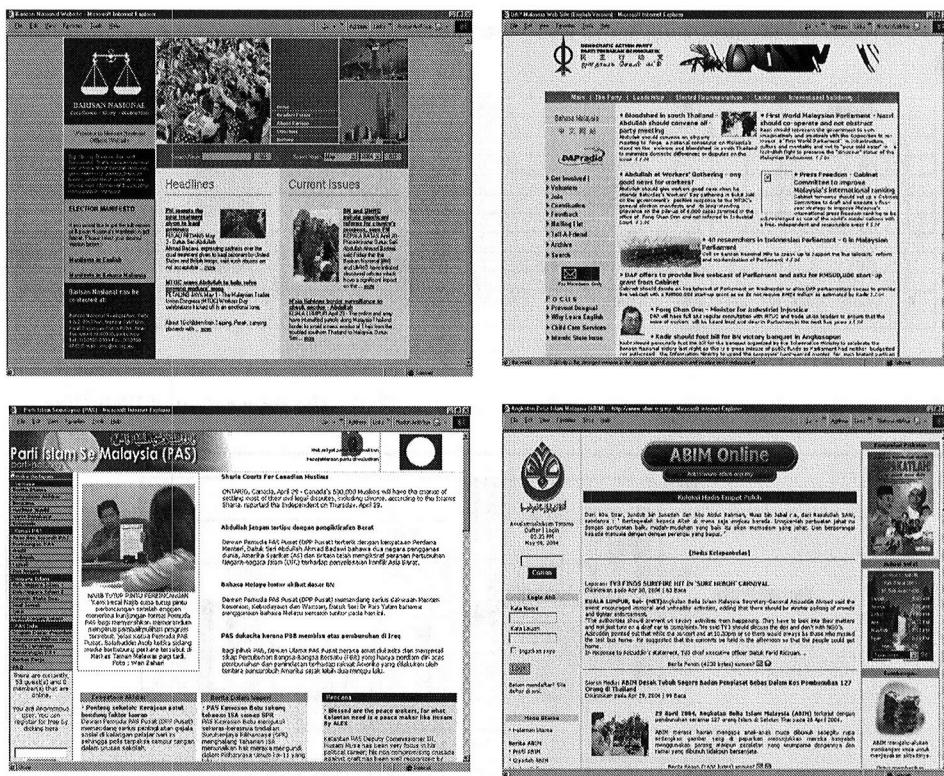
Step 8. Monitor Progress

After the redesign process (if necessary), the next step is to monitor the progress of the new version of the Web site. Several measures can be used for this, such as counting page hits, tracking user logs, and identifying the level of sales volume (in the case of e-commerce sites). After several months, Web benchmarking should be repeated to track progress as compared to others in similar fields or businesses.

FRAMEWORK TESTING

Once completed, the benchmarking framework was tested for its applicability and practicality. The main purposes of the benchmarking are threefold: (1) to test the suitability of the criteria in terms of wordings and terminology used, (2) to test the practicality of the eight benchmarking steps and proposed calculation methods, and (3) to identify the level of usability of major political Web sites in Malaysia. The benchmark only focussed on the benchmarking of Web usability that covers the key areas of Screen Appearance, Consistency, Navigation, Media Use, Interactivity, and Content. The list of the objective criteria used as the benchmarking metrics is shown in Tables 4 through 10. This benchmark could be considered as an external benchmarking because it involved comparisons between different political Web sites in Malaysia. Four major political Web sites were selected

Figure 2. Sample Web pages of four selected political Web sites used in the benchmarking: The National Front Party (<http://www.bn.org.my>), The Malaysian Pan Islamic Party (<http://www.parti-pas.org>), The Democratic Action Party (<http://www.malaysia.net/dap>), and The Islamic Youth Movement (<http://www.abim.org.my>)



— National Front Party (BN), Malaysian Pan Islamic Party (PAS), Democratic Action Party (DAP), and Islamic Youth Movement (ABIM) (see sample of screenshots in Figure 2).

Two evaluators (expert Internet users) were invited to participate. They conducted the evaluation in a room with two computers with the specifications as presented in Table 15.

The evaluators were briefed on the purposes of the benchmarking and what they were supposed to do. Benchmarking forms were supplied to the evaluators be-

fore they started the benchmarking. Using the forms, the evaluators then performed the benchmarking for about three hours on the selected Web sites. After the benchmarking, the forms were collected from the evaluators. The number of criteria *existence* and *non-existence* were calculated and summarised as presented in Tables 16(a) and (b). The percentage usability index for all four Web sites together with a bar chart were also presented.

The results show that all four political Web sites have very good design in terms of screen appearance. In general, design-

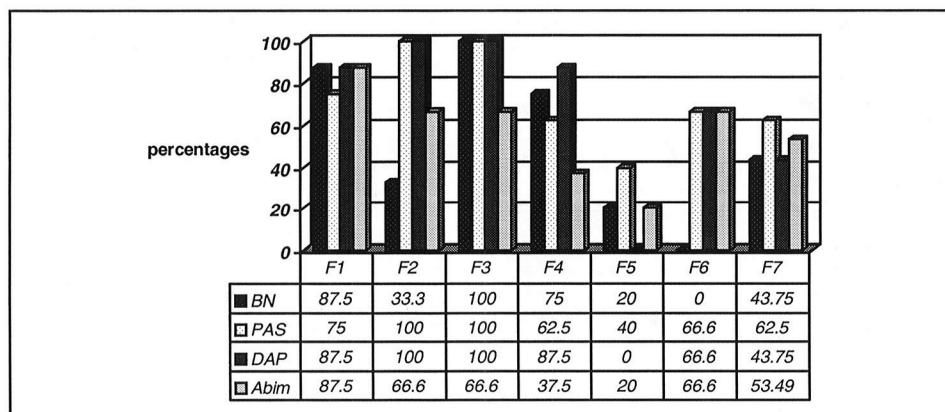
Table 15. Computer specifications for Web benchmarking

	<i>Processor</i>	<i>RAM</i>	<i>Internet access</i>	<i>Internet browser</i>	<i>Screen resolution</i>
Computer 1	Pentium 1	32 Mb	33 Kbps Modem	Netscape 4.5 & Explorer 5.0	640 X 480 pixels
Computer 2	Pentium 2	128 Mb	56 Kbps Modem	Netscape 6.2 & Explorer 6.0	800 X 600 pixels

Table 16(a). The benchmarking score and percentage Web usability index for the selected Web sites

Factors	BN			PAS			DAP			ABIM		
	YES	NO	NA	YES	NO	NA	YES	NO	NA	YES	NO	NA
F1 (/8)	7	1		6	2		7	1		7	1	
F2 (/3)	1	2		3	0		3	0		2	1	
F3 (/3)	3	0		3	0		3	0		2	1	
F4 (/8)	6	2		5	3		7	1		3	5	
F5 (/5)	1	1	3	2	2	1	0	2	3	1	1	3
F6 (/3)	0	3		2	1		2	1		2	1	
F7 (/16)	7	9		10	6		7	9		6	10	
Total (/46)	25	18	3	31	12	1	29	14	3	23	20	3

% Usability Index for: BN=25/(46- 3)X100=58.14%, PAS=31/(46-1)X100=68.89%, DAP=29/(46-3)X100=67.44%, ABIM=27/(46-3)X100=53.49%

Table 16(b). Bar chart of the benchmarking score (in %) for the four selected Web sites

ers of these sites utilised proper colour, text, titles, headings, and layout. In terms of consistency, Web sites belonging to PAS and DAP were very consistent in all three aspects of page layout, use of text, and navigational aids. The other two Web sites (i.e., BN and ABIM), however, suffered from page layout inconsistency such as placement of content display and banners. Apart from page layout, BN's Web site also has inconsistent use of text in terms of its types,

font size, and colour. Three Web sites — BN, PAS, and DAP — were highly accessible in both aspects of display compatibility and searching facility. ABIM's Web site, however, did not provide any searching function for better accessibility.

DAP's Web site has the highest level of usability in navigation category. BN and PAS also rated well in this category. However, ABIM's Web site had major navigation problems, including a few broken links,

long page scrolling, and unavailability of a site map. In the media use category, most Web sites did not utilise continuous media in presenting information. All sites also failed to properly use static media where graphics, logos, and pictures were not labelled. However, most Web sites were rated better in terms of interactivity. Although features for entertainment were not available, Web sites belonging to PAS, DAP, and ABIM provided all features for users' feedback and discussions. BN's Web site, on the other hand, had very severe interactivity problems where all three criteria were absent. When benchmarking content criteria, PAS and ABIM performed better than BN and DAP. Both had a wider scope of contents, especially those related to content authority and linkages. Despite performing slightly worse than the others in most aspects, BN's Web site scored high for content in terms of authority and reliability. On the other hand, DAP's Web site suffers severe linkages and text quality problems.

In practise, Web benchmarking is normally performed by an organisation by comparing its Web site with its competitors. Therefore, the result can be used to make changes for better Web sites in terms of usability. However, the benchmarking in this research was only to test the applicability of the framework and not being performed on behalf of any particular organisation. Hence, steps 7 and 8 are not contextually applicable. Nonetheless, the results of the benchmarking revealed some usability problems faced by all organisations as described in step 6. In general, the usability level of the Web site belonging to PAS has the highest level of usability with 68.89%, followed by DAP, BN, and ABIM with 67.44%, 58.14%, and 53.49%, respectively. The results also provide ideas to designers of all these Web sites, particularly

BN and DAP on areas that need to be concentrated on in the redesign of their sites.

Outcomes of the Framework Testing

The benchmarking was conducted successfully with satisfactory results. The benchmarking processes or steps were easily followed and executed. Good feedback was obtained from the evaluators. The criteria used were easily understood and evaluated. The number of criteria for all categories was also considered adequate. Nonetheless, after the testing, several issues were noted to improve the applicability of the framework, which include:

1. The whole process of performing the benchmarking in step 5 was very time consuming, particularly when the evaluators had to go through every Web page in the site to assess a criterion (e.g., clear title for each Web page). Two solutions are recommended to minimise this problem as follows:
 - During step 4, select more evaluators (e.g., one evaluator for each SCANMIC category).
 - During step 5 (i.e., perform the benchmark), instead of evaluating all Web pages, allow evaluators to test parts of the Web site (e.g., if the site has five sub-categories, probably evaluating at least two pages for each category is adequate).
2. The outcome of the testing also suggests that the benchmarking evaluation method needs to be expanded, particularly for step 7 (i.e., redesign). In addition to relying on the results of step 6 (i.e., analyse data and determine gap) to redesign the Web site, other evaluation methods (e.g., expert reviews) could also be used, par-

ticularly those that deal with the assessment of the subjective criteria. Therefore, it should be mentioned in step 7 that the results of the benchmarking, together with the results of other assessment methods (e.g., expert review, interview, and user observation), should be utilised in the redesign process.

The testing also reveals the need for individuals, organisations, and the government who are involved in political Web site design and content development to put extra effort in raising the usability level of their Web sites. The results from the testing shows that all major political Web sites in Malaysia still suffer some severe usability flaws that need to be tackled immediately. This type of Web sites plays a very important role in bridging the relationship between the government, business, and the public at large (consumers).

CONCLUSIONS AND SUGGESTIONS FOR FUTURE STUDIES

There are many ways that companies can assess their Web presence including usability testing, questionnaire survey, interview, and expert review. Each method has its advantages and is being used to achieve certain objectives. Using only one method is not adequate to assess the quality of one's Web site. Thus, combining several approaches in Web evaluation would produce better results. This study provides an alternative approach to Web evaluation through the benchmarking approach. It highlights the need for Web developers to benchmark the usability and content usefulness of their sites against their competitors. Apart from being able to identify their sites' weaknesses, the main advantage of

this method is that Web developers would also be able to find out their competitors' strengths and then redesign their own Web sites for better usability.

The metrics used for our benchmarking are derived from rigorous analysis of the literature and expert verification. The groupings are also refined based on comments and recommendation from experts. Only key generic criteria are used so that they are applicable to all types of Web sites. Furthermore, only objective criteria are used for this framework. Based on our empirical study, the proposed framework can be applied to the real-world situation. In addition, the outcome of the framework testing allows a refinement to the framework. Most importantly, the cyclical nature of the benchmarking processes as proposed in the framework is very suitable for Web evaluation due to the changing nature of Web technology and user requirements. The proposed benchmarking approach also has some advantages compared to other methodologies, such as those developed by Misic and Johnson (1999), Barnes and Vidgen (2002), and Ivory-Ndiaye (2003). Its main advantage is in terms of coverage, whereby our approach covers wider usability issues including media use, accessibility, and content. Additionally, the proposed approach does not require many Web users or experts to benchmark a particular site, which would certainly speed up the benchmarking process. The ease-of-use of the calculation steps is also an added advantage.

However, there are a few points that are worth mentioning for further research. First, the framework was only tested on political Web sites. Further efforts should be made by testing the framework on other type of Web sites including e-commerce sites. The results could then be used to strengthen the framework. Second, this

study only deals with objective Web criteria. Although subjective criteria were identified during the early stage, no suggestion was given on how to tackle these criteria. Further studies are needed on how these criteria could be used in Web evaluation. Third, since the proposed framework only considers the objective measures, further work can be performed including automating the benchmarking process. Fourth, further study is also needed to identify the relative importance of each criterion within the same category or factor. Finally, this study only deals with the issue of Web usability and content usefulness. Many other factors, including technological, cultural, social, economic, and legal factors, should also be considered in future work.

REFERENCES

- Anderson, B. & Pettersen, P.G. (1996). *The benchmarking book, step-by-step instruction*. UK: Chapman and Hall.
- Barnes, S. & Vidgen, R. (2002). An integrative approach to the assessment of e-commerce quality. *Journal of Electronic Commerce Research*, 3(3), 114-127.
- Bendell, T., Boulter, L., & Kelly, J. (1993). *Benchmarking For competitive advantage*. UK: Pitman Publishing.
- Bramham, J. (1997). *Benchmarking for people managers*. UK: Cromwell Press.
- Chang, R.Y. & Kelly, P.K. (1995). *Improving through benchmarking*. London: Kogan Page.
- Codling, S. (1992). *Best practice benchmarking, the management guide to successful implementation*. UK: Industrial Newsletters Ltd.
- Elmuti, D. (1998). The perceived impact of the benchmarking process on organisational effectiveness. *Production and Inventory Management Journal*, (3rd Quarter), 39.
- Government Centre for Information Systems. (1995). Improving for money from IS/IT service provision. *Benchmarking IS/IT*. Norwich, UK: HMSO Publications.
- IBM. (2000). *IBM Web design guideline*. Retrieved January 3, 2000: http://www-3.ibm.com/ibm/easy/eou_ext.nsf/publish/572
- Ivory-Ndiaye, Y.M. (2003). An automated approach to Web evaluation. *Journal of Digital Information Management*, 1(3), 75-102.
- Lu, M. & Yeong, W. (1998). A framework for effective commercial Web application development. *Internet Research Journal*, 8(2), 166-173.
- Lynch, P.J. & Horton, S. (1999). Interface design for WWW Web style guide. *Yale Style Manual*. Retrieved December 25, 1999: <http://info.med.yale.edu/caim/manual/interface.html>
- Misic, M.M. & Johnson, K. (1999). Benchmarking: A toll for Web site evaluation and improvement. *Internet Research: Electronic Networking Applications and Policy*, 9(5), 383-392.
- Morkes, J. & Nielsen, J. (1998). *Applying writing guidelines to Web pages*. Retrieved November 25, 1999: <http://www.useit.com>
- NetMechanic. (2000). *Netmechanic*. Retrieved July 20, 2000: <http://www.netmechanic.com>
- Nielsen, J. (1993). *Usability engineering*. San Diego: Academic Press.
- Nielsen, J. (1997a). Changes in Web usability since 1994. *Jacob Nielsen's Alertbox*. Retrieved December 23, 1999: <http://www.useit.com/alertbox/9712a.html>

- Nielsen, J. (1997b). Report from 1994 Web usability study. *Papers and Essays*. Retrieved January 15, 2000: http://www.usability.com/papers/1994_Web_usability_report.html
- Perlman, G. (2001). *Suggested readings in human-computer interaction (HCI), user interface (UI) development, & human factors (HF)*. Retrieved June 14, 2002: <http://www.hcibib.org/readings.html>
- Powell, A.T. (2000). *Web design, the complete reference*. USA: Osbourne/McGraw-Hill.
- Seminario, M. (1998). Study: One in three experienced surfers find online shopping difficult. *ZDNet*. Retrieved June 8, 2000: <http://www.zdnet.com/intweek/quickpoll/981007/981007b.html>
- Shackel, B. (1991). Usability^¾context, framework, design, and evaluation. In B. Shackel & S. Richardson (Eds.), *Human factors for informatics usability* (pp. 21-38). Cambridge: Cambridge University Press.
- Shirley, H. (1999). *Effective electronic training, designing electronic materials: Articles and papers*. Retrieved November 9, 1999: <http://www.rockley.com/designin.htm>
- Shneiderman, B. (1997). Designing information-abundant Web sites: Issues and recommendations. *International Journal of Human Computer Studies*, 47, 5-29.
- Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd edition). USA: Addison Wesley Longman.
- Simeon, R. (1999). Evaluating domestic and international Web site strategies. *Internet Research: Electronic Networking and Policy*, 9(4), 297-308.
- Smith, P.A., Newman, I.A., & Parks, L.M. (1997). Virtual hierarchy and virtual networks: Some lessons from hypermedia usability research applied to WWW. *International Journal of Human-Computer Studies*, 47(1), 67-95.
- Sun Microsystems. (1999). *Writing for the Web guide*. Retrieved November 10, 1999: <http://www.sun.com>
- Tauscher, L. & Greenberg, S. (1997). How people revisit Web pages. *International Journal of Human-Computer Studies*, 47(1), 97-137.
- UsableWeb.com. (2002). *Keith Instone's Web site*. Retrieved July 10, 2002: <http://usableWeb.com>
- Useit.com.(2002). *Jacob Nielsen's Useit*. Retrieved July 10, 2002: <http://ww.useit.com>
- W3C. (2002). *Web content accessibility guidelines 1.0*. Retrieved November 14, 2003: <http://www.w3.org/WAI/EO/Drafts/impl/eval/Overview.htm>
- Web Workshop. (1999). *Improving Web site usability and appeal*. Retrieved December 12, 1999: http://msdn.microsoft.com/workshop/management/planning/improving_siteuser.asp
- Zhang, P. & Dran G.M. (2000). Satisfiers and dissatisfiers: A two-factor model for Web site design and evaluation. *Journal of the American Society for Information Science*, 51(14), 1253-1268.
- Zimmerman, E.D., Muraski, M., Palmquist, M., Estes, M., McClintoch, C., & Bilsing, L. (1998). *Examining WWW designs: Lessons from pilot studies*. Retrieved October 24, 1999: <http://www.microsoft.com/usability/Webconf/zimmerman.htm>

Associate Professor Norshuhada Shiratuddin holds a PhD in computer and information sciences from the University of Strathclyde, Glasgow, Scotland; a master's degree in information technology from the University of Nottingham; and a BSc from UMIST, Manchester, UK. She is currently attached to the Universiti Utara Malaysia as one of the faculty postgraduate coordinators. Her research interests include software engineering, electronic books/multimedia design and development, and Web publishing. She is actively publishing her works in international journals and proceedings (for the past six years she has published about 40 articles), books and monographs, particularly research findings on digital content in education.

Feng Li (fengli@ncl.ac.uk) is chair of e-business at the University of Newcastle upon Tyne Business School, UK. His research has focused on the interactions between information systems and emerging strategies, business models, and organizational designs. He is the author of two books and numerous journal articles. Professor Li speaks regularly at international conferences and to business executives from both the private and public sectors. Professor Li is a member of several programs on ICTs, e-commerce/e-business, supply chain/value chain, and virtual teams. He has worked closely with companies in banking, telecommunications, manufacturing, retailing and electronics as well as the public sectors. He is the e-business SIG chair in British Academy of Management (BAM), and is the winner of the Blackwell Prize for E-Business and Technology Management at BAM2002. His recent work on Internet banking and telecommunication pricing models and value networks has been extensively reported by the media.