DESIGN METRICS AND THE ADAPTATION OF WEB-PAGE CONTENT CHUNKS FOR PDAs

S.D. Scott[a] and Y.H. Koh[b]

[a]BTexact Technologies, Malaysia. email: simon@apiit.edu.my
Faculty of Engineering, Malaysia,
[b]Multimedia University, email: v9111316@mmu.edu.my

***Abstract* - The majority of web-pages are unsuitable for viewing on PDAs, WAP phones and similar devices without first being adapted. However, little empirical work has been done on what actually constitutes a good PDA or WAP web-page. This paper ranks a number of PDA web-pages from different categories empirically and correlates the results against the design metrics present. The findings are then compared against a similar set of experiments for PC web-pages. The results of this comparison suggest that, as well as omitting, summarizing and converting individual multimedia objects in the web-page to a less resource intensive form, the design metrics need to be changed during adaptation to enhance the presentation of web-content on non-PC devices. The paper concludes by investigating the effect of applying some suitable changes to the design metrics on web-page content chunks, which form the basic units in automatic content adaptation systems.**

***Keywords:* - Content Adaptation, Personalization, Design Metrics.**

1. INTRODUCTION

The majority of web-sites are built under the assumption that all users are accessing the site from a PC. In other words, these web-sites are written in markup languages that presume devices have similar display sizes, memory and software capabilities and are largely oblivious to the clients' available network bandwidth and perceived network latency. As a result of this assumed device heterogeneity, clients may receive content that they cannot store, that they cannot display, that violates the desires of the user, or takes too long to convey over the network to the client device. This state of affairs has been known for some time (Katz, 1994), but is becoming more critical with the increasing popularity of WAP and recently PDA devices. To avoid this potential wastage of bandwidth and processing resources the content transmitted to the client from a web-site needs to be adapted to both the device and user.

There are two main approaches to adapting web-content. The first approach is to create different versions of the same content, store the different versions on the server and use the device characteristics to decide which version of the content to send to the client. This is usually achieved using the eXtensible mark up language (XML) and the associated Stylesheet language (XSL), see examples (Coninx *et al.*, 2003; Kirda *et al.*, 2001; Korolev & Joshi, 2001). The problem of this approach is that it involves a lot of repetitive work on the part of the web-site administrator in editing multimedia objects down to different resolutions and constructing multiple layout templates for the objects as well as continual maintenance of the above as new technologies and content becomes available.

The second approach is to dynamically alter the content retrieved by the web-server before it is displayed, usually through the use of proxy servers. A number of such automatic content adaptation systems have been developed by various groups (Banerjee *et al.*, 2003; Bharadvaj *et al.*, 1998; Bickmore *et al.*, 1999; Ma *et al.*, 2000). In these systems, the individual multimedia objects in a web-page are adapted through omission, summarization or conversion to a less resource intensive form. As the size of the display changes, the layout of the multimedia objects also needs to change. Changing the layout requires semantic information to be extracted from the web-site; in other words the system needs to know how the different elements in the page are related, and their functionality. Work in this area includes detecting the purpose of individual multimedia objects (Paek & Smith, 1998), determining how multimedia objects in pages are related to one another (Chen *et al.*, 2001; Soumen *et al.*, 2001), and extraction of specific functional information from the entire web-site (Schilit *et al.*, 2001).

Central to adaptation of web-content is the question of what makes a good web-page on an arbitrary device. There are numerous detailed usability guidelines available for desktop PC based web-page design, however these have largely been drawn up from the personal experience of the various authors and there is little consistency between them (Ratner *et al.*, 1996). Further, at present, there are no comprehensive design guidelines available for devices other than desktop PCs. Of those design guidelines that do exist, a few of them deal with issues such as text layout and pagination (Giller *et al.*, 2003), whilst the majority of them deal with the limitations of the devices (Schmidt *et al.*, 2000).

Design metrics are measures relating to composition (e.g. word count, link count), formatting (e.g. emphasized text, positioning) and other general characteristics (e.g. total bytes) that influence usability (Ivory *et al.*, 2000). A quantitative analysis of a large collection of web-sites has been performed at Berkeley by Ivory, Hearst and Sinha, who introduced a methodology whereby a number of computed design metrics are used to predict scores assigned by expert judges (Ivory *et al.*, 2001; Ivory *et al.*, 2002). This methodology allowed them a 94% success rate in predicting whether a web-page would be ranked as good, average or bad. Further, they found that both web-pages with different functions and web-pages containing content relating to different subject matter have different characteristics. For example, home pages have different design metrics compared to menu pages and pages with finance content have different metrics compared to pages with educational content. This is not to say that form is more important than content; rather that good form does increase the usability of web-pages and that those authors who make more of an effort to provide good content are more likely to ensure it is presented in an appropriate form.

2. PDA DESIGN METRICS

2.1 *Experiment*

To investigate PDA design metrics, the following four categories of PDA web-pages were defined, namely: i) home pages - the default page on entering a web-site, often a logo, description and a set of links; ii) menu pages - pages containing lists of links, usually with accompanying description; iii) content pages - pages containing an article with least 300 words; and iv) data entry pages - pages containing a set of forms for data capture. An example web-page from each of these categories is shown in Figure 1. Although these categories are not exhaustive and some PDA web-pages fall into more than one category, they were selected since a plausible strategy for content adaptation plausible is to construct pages based on function (Chen *et al.*, 2001). The question of whether PDA web-pages containing content on different topics (e.g. education, commerce) had different metrics was not investigated since the number of PDA web-pages in the public domain at present is somewhat limited.

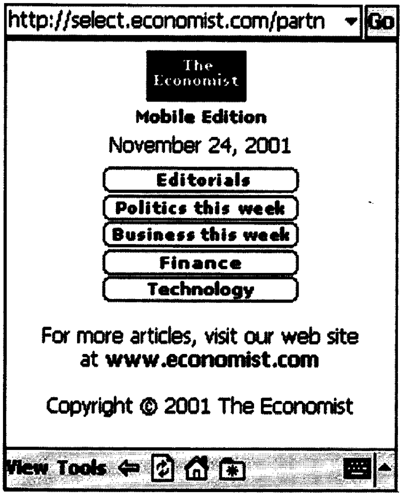
The experiment utilized a Compaq iPAQ pocket PC (screen size 240x320) with Microsoft's Pocket IE 3.0 installed. The first thirty suitable commercial PDA-web pages located for each category were downloaded and stored into the local cache of the iPAQ. Eleven IT proficient subjects were then asked to rate the pages, viewed on the iPAQ, giving a single mark between zero (unusable) and ten (excellent). In rating the page, the subjects were asked to consider the following: i) content, ii) structure and navigation, iii) visual design, iv) functionality, v) interactivity and vi) overall design. These factors are the same as those in the 2000 Webby awards (The Webby Awards, 2003) and the Berkeley experiments. A single mark was given rather than six individual marks since it has already been demonstrated that 90% of the variance in the 2000 Webby awards can be accounted for by a single variable (Ivory *et al.*, 2001). The users were asked to view all the pages before rating them so as to prevent biases in the pages viewed first.

For each category, the thirty pages were split into good, average and bad sets (ten pages in each set) based on the probability that they were above or below the rest of the pages in the same category using a 1-sided t-test. This technique allowed a decision to be made between pages that had the same number of total marks. For example, the 10^{th} and 11^{th} highest rated home pages both had the same number of marks but the 10th highest rated home page had a probability of 0.9 that it was better than the other home pages while the 11th highest rated page only had a probability of 0.7. To give an idea of the range, the probability that the highest rated home page (MSN Mobile) was better than the other home pages was found to be over 0.99999.

For the purpose of comparison with work done at Berkeley on PC web-page metrics, the same set of 11 metrics were employed (Ivory *et al.*, 2001), see Table 1 below. The values of these of metrics were calculated for all the PDA web-pages. The statistics of the metric values for the good and bad sets of pages, along with their significance are shown in Table 2.

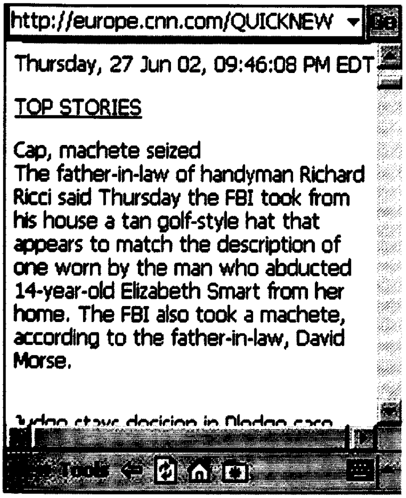
Table 1: Web-page metrics (Ivory *et al.*, 2001).

Metric	Description
Word Count	Total words on the page
Body Text	Percentage of words that are body text vs display text (i.e. headers)
Emphasized Body Text	Percentage of body text that is emphasized (e.g. bold, capitalized or near !'s)
Text positioning count	Changes in text position from flush left
Text cluster count	Text areas highlighted with color, bordered regions, rules or lists
Link count	Total links on the page
Page Size	Total bytes for the page as well as elements, graphics and stylesheets
Graphic	Percentage of page bytes that are for graphics
Graphics Count	Total graphics on the page (not including graphics specified in scripts, applets or objects)
Color Count	Total colors employed
Font Count	Total fonts employed (i.e., face + size + bold + italic)

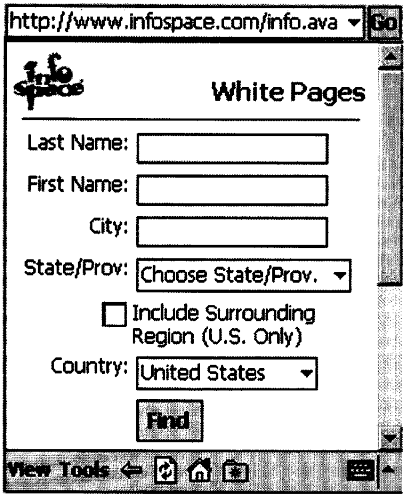


Home Page

Menu Page



Content Page



Data Entry Page

Figure 1: An example page from each of the PDA web-page categories.

Table 2: Means and standard deviations for the good (G) and bad (B) sets of pages in the different function categories. The table also contains t-test results (two-tailed significance) for each metric. The bold text indicates the significant results.

Metric	Home Pages			Menu Pages			Content Pages			Data Entry Pages		
	G	B	Sig.	G	B	Sig.	G	B	Sig.	G	B	Sig.
Word Count	44 (41)	87 (212)	.559	160 (118)	82 (31)	.091	814 (414)	855 (302)	.790	54 (57)	67 (109)	0.741
Body Text (%)	40 (29)	26 (38)	.439	**47 (32)**	**9 (17)**	**.008**	95 (4)	97 (2)	.392	**27 (27)**	**55 (29)**	**0.025**
Emphasized (%)	0.6 (1.8)	0.3 (1.1)	.727	4.0 (12.0)	0.7 (2.3)	.466	9.5 (26.6)	0.2 (0.5)	.299	9.5 (15.8)	14.7 (31.3)	0.656
Text positioning	4.1 (2.8)	2.9 (2.0)	.343	1.8 (1.0)	1.8 (1.0)	.000	2.0 (0.9)	2.8 (2.3)	.351	2.1 (1.5)	2.8 (1.9)	0.511
Text clusters	4.9 (2.5)	2.6 (1.3)	.055	**8.4 (3.5)**	**4.8 (2.9)**	**.042**	7.2 (2.6)	5.5 (2.0)	.209	5.0 (2.8)	4.4 (2.2)	0.658
Link count	**10.5 (7.0)**	**5.1 (3.5)**	**.048**	18.2 (11.3)	17.6 (8.4)	.845	5.1 (4.1)	3.8 (2.7)	.474	5.1 (4.1)	5.8 (7.5)	0.805
Page Size (kB)	5.6 (4.9)	2.6 (1.5)	.110	5.6 (2.4)	4.2 (3.9)	.378	**9.8 (2.2)**	**6.8 (2.0)**	**.015**	**8.2 (6.0)**	**3.4 (2.1)**	**0.042**
Graphic (%)	54 (23)	47 (27)	.588	32 (27)	29 (29)	.820	**29 (24)**	**12 (14)**	**.029**	14 (14)	12 (24)	0.846
Graphics Count	4.2 (4.5)	1.5 (1.0)	.081	2.9 (3.4)	1.6 (3.0)	.433	**2.2 (1.4)**	**0.8 (0.1)**	**.013**	1.5 (2.3)	0.5 (0.5)	0.186
Color Count	**2.0 (0.5)**	**1.6 (0.5)**	**.037**	2.2 (0.3)	2.0 (0.0)	.343	1.9 (0.6)	1.7 (0.5)	.443	1.7 (0.5)	1.4 (0.5)	0.193
Font Count	**3.7 (1.5)**	**2.3 (1.3)**	**.050**	**4.6 (1.4)**	**2.8 (0.9)**	**.004**	4.7 (1.4)	4.4 (1.6)	.647	3.2 (0.8)	3.3 (1.6)	0.876

2.2 *Discussion*

Some general patterns are apparent from Table 2:

- Good pages are larger in size (bytes) than bad pages. This is mainly due to good pages having more graphics than bad pages.
- Good pages have more colors (text) than bad pages, though this difference was only significant in the home page category.
- Good pages tend to have more text clusters than bad pages, though this difference was only significant in the menu page category. This metric can be equated with clear page layout.
- Good pages tend to use more fonts than bad pages, this is a significant factor in both the home and menu page categories.

The individual significant results are commented on below:

- The difference in the average amount of body text in good and bad menu pages may have been because a larger amount of text facilitates a clearer description of what the links contain. However, this result might also have been a consequence of the nature of the task (browsing) in the experiment. If the users were already familiar with the type of content provided by the link then perhaps the result would not have been so significant.
- In the case of data entry pages, it seems that too much body text can spoil the layout of the page, offsetting any benefits it might have in clarifying the nature of the data to be entered.
- Good home pages have high link counts. This is probably because the users expect home pages to be useful rather than decorative. PDA home pages that consisted of a single link / graphic received low ratings in the experiment.
- Good data entry pages are large; note that this is not due to graphics but because items like drop down menus take up memory and are rated highly since they make inputting data easier and allow a cleaner layout in general. To confirm this, the number of items (text boxes, list boxes, buttons, etc) in each page was calculated but no significant difference was found between the good and bad sets of data entry pages.
- Good content pages are large, probably because graphics usual add to the acceptability of the pages. Note that without graphics the values of the mean and standard deviation (in brackets) for the size of the good and bad sets of content pages were 6.8 (2.3) and 6.0 (1.8) respectively, which was not found to be significant.

Examination of the sets of pages suggested a number of other metrics that could be used to distinguish between good and bad pages in the different categories (Ivory *et al.*, 2000):

- Good home pages utilize more of the available screen space. This observation was confirmed by calculating the percentage of white space in the home pages. The values of the mean and standard deviation (in brackets) for the percentage of white space in the good and bad sets of home pages were 10 (15) and 35 (27) respectively, which had a two-sided t-test significance of 0.02.
- Good menu pages have a reasonably consistent number of words per link cluster. The values of the mean and standard deviation (in brackets) for the average number of words per link cluster for the good and bad sets were 25 (9) and 41 (33) respectively. However the difference was not significant (0.09 for a two-sided t-test).

- Good content pages use headers (usually a bold font) to introduce new themes or split the articles into subsections. Half of the good content pages used such headers whilst none of the bad pages did so, which is significant at the 5% level.
- Good content pages use larger fonts. The values of the mean and standard deviation (in brackets) for the font size of the good and bad content pages were 11.4 (1.0) and 10.5 (2.1) respectively. However, this difference was not significant.

2.3 *Comparison with PC Web-Page Metrics*

In the Berkeley experiments on PC web-page metrics it was found that the word count significantly correlated with all the metrics except emphasized text, and so the pages were sub-divided into three groups depending on their word count, namely low, medium and high word count (Ivory *et al.*, 2001). In the case of the PDA web-pages it was found that the word count significantly correlated with the body text percentage, cluster count, page size, graphic percentage (negative correlation) and font count. There are two possibilities as to why the word count in PDA web-pages correlated with less metrics than in PC web-pages. The first is that the PDA-pages are fundamentally different from PC pages due to the limited display, for example text positioning count might be less relevant since there is much less scope for text positioning in PDA web-pages. The second reason is that the PC web-page data set contained a much larger percentage of level two pages (defined as two mouse clicks away from the home page of a web-site) than the PDA web-page data set, and so contains a much smaller percentage from home and data entry web-pages.

The PDA web-pages used in the above experiment were split into three groups based on the word count in the pages. The high word count group (avg. word count = 808) mainly consisted of pages from the content category; the medium word count group (avg. word count = 196) mainly consisted of pages from the menu category; and the low word count group (avg. word count = 37) mainly consisted of pages from the home and data entry categories. For each word count group the PDA web-pages were split into good, average and bad sets based on the marks awarded to them. The statistics of the metric values for the good and bad sets of pages, along with their significance are shown in Table 3.

Comparing our statistics in with those in the Berkeley experiment leads to the following observations:

- In all three PDA web-page word count groups, the good sets of pages had more emphasized body text than the bad sets of pages. This is contrary to PC web-pages where the good sets of pages had less emphasized body text than bad sets.

Table 3: Means and standard deviations for the good (G) and bad (B) sets of pages with different word counts. The table also contains t-test results (2-tailed significance) for each metric. The bold text indicates the significant results.

Metric	Low Word Count			Medium Word Count			High Word Count		
	G	B	Sig.	G	B	Sig.	G	B	Sig.
Word Count	36 (24)	38 (29)	0.782	173 (79)	204 (85)	0.304	847 (378)	761 (225)	0.530
Body Text (%)	33 (29)	21 (30)	0.214	48 (31)	50 (32)	0.888	92 (9)	87 (29)	0.622
Emphasized (%)	**2.1 (4.6)**	**0.2 (1.0)**	**0.046**	10.8 (18.0)	2.4 (5.5)	0.12	10.0 (26.5)	0.5 (1.1)	0.281
Text positioning	2.6 (1.7)	2.8 (2.0)	0.714	1.7 (1.1)	3.0 (1.8)	0.146	2.6 (2.1)	2.8 (2.2)	0.862
Text clusters	4.4 (2.3)	3.5 (2.3)	0.220	**8.1 (2.5)**	**5.4 (2.3)**	**0.046**	9.3 (4.7)	6.3 (2.1)	0.101
Link count	8.2 (9.8)	9.8 (9.3)	0.557	11.9 (10.3)	14.2 (5.2)	0.530	7.1 (4.8)	9.8 (18.1)	0.685
Page Size (kB)	**6.4 (4.6)**	**2.3 (1.0)**	**0.001**	7.3 (5.0)	5.6 (3.8)	0.312	**9.2 (2.8)**	**6.7 (2.0)**	**0.026**
Graphic (%)	26 (25)	32 (31)	0.580	29 (26)	25 (21)	0.801	21 (23)	14 (14)	0.307
Graphics Count	2.2 (3.1)	0.9 (1.0)	0.123	4.2 (4.4)	2.8 (3.1)	0.433	**1.9 (1.5)**	**1.0 (0.7)**	**0.041**
Color Count	1.9 (0.5)	1.7 (0.5)	0.187	2.0 (0.7)	1.9 (0.3)	0.678	1.9 (0.7)	1.8 (0.4)	0.758
Font Count	**3.6 (1.3)**	**2.5 (1.0)**	**0.010**	4.5 (1.6)	3.7 (1.2)	0.168	5.0 (1.5)	4.4 (1.6)	0.515

- In all three PDA web-page word count groups, the good sets of pages had a higher graphics count than the bad sets of pages. Again this is contrary to PC web-pages where the good sets of pages had a lower graphics count than bad sets. Since the percentage of the page size given over to graphics in PDA web-page was not a significant factor, using large numbers of small graphics seems to lead to better PDA web-pages than small numbers of large graphics (in terms of bytes required).
- For both PDA and PC web-pages, the good sets had more text clusters and more fonts than bad sets regardless of overall word count. Further the number of fonts and clusters in PDA and PC web-pages was comparable.
- PDA web-pages use a lower number of colors (usually just two colors) compared to PC web-pages (seven colors on average). Further, the number of colors does not increase with increasing word count in PDA web-pages, unlike in PC web-pages.

- PDA web-pages have less links (avg. ≈ 10) compared to PC web-pages (avg. ≈ 40). This is probably because PC web-pages are usually made up of a number of frames, some of which contain lists of links for easy navigation.
- For PDA web-pages, bigger is better, which is not always the case for PC web-pages. However, given that in the PDA experiment the pages were loaded from a cache, whilst in the Berkeley experiment they were downloaded from the Internet, one should be careful about reading anything into this observation.

The above results show that PDA web-pages have different design metrics compared to PC web-pages. This is an important result for automatic content adaptation since it means that specific information needs to be added to, or removed from, individual multimedia objects to improve their presentation. For example, more text needs to be emphasized and the less colors used. In other words, systems that adapt content solely by omission, summarization or conversion to a less resource intensive form are likely to produce less acceptable results than those that also adjust the design metrics of the objects.

3. LAYOUT SCHEME

Web pages can be split up into a number of content chunks. These chunks contain a set of multimedia objects that relate to one particular area of interest or task. If a basic object is defined as one that contains a single multimedia element (for example an image or a body of text), and a composite object is defined as a set of objects (basic or composite) that perform some certain functions together (Yang & Zhang, 2001), then a chunk is itself a high-level composite object. This paper is concerned with the adaptation of individual content chucks (identified manually), rather than whole web-pages, since such chunks form the natural units of a web-page when is broken up for such a purpose (Chen *et al.*, 2001). In the adaptation process outline below, a layout scheme is first applied that re-arranges and resizes the individual multimedia objects in the chunk so that the width constraints of the device are satisfied. Second, appropriate sets of rules for altering the design metrics are applied. Although changes due to the layout scheme and the metric rules can overlap, they are kept separate since the layout scheme is general, but the metric rules will be device specific.

A number of adaptation rules for content chunks have been outlined elsewhere (Chen *et al.*, 2001; Yang & Zhang, 2001), however these are concerned with how content chunks should be split over a number of pages rather than providing an algorithm for re-arranging the layout of the content chunks themselves to fit an arbitrary screen width. To understand what constitutes a good layout, a multimedia designer was commissioned to re-arrange and alter the formatting design metrics of thirty content chunks randomly selected from various commercial Internet sites, so that they were suitable for viewing on a Compaq iPAQ pocket PC (screen size 240x320). By viewing the before and after arrangement of the composite objects in the chunks a simple scheme, see Figure 2, was determined.

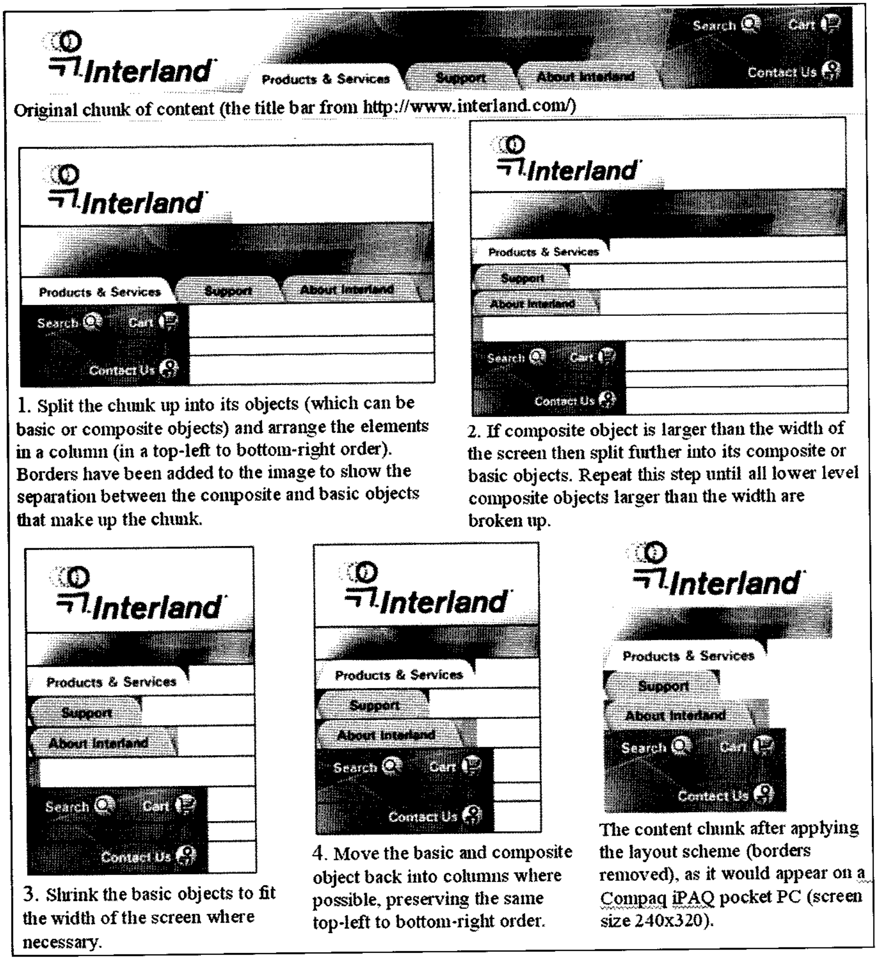


Figure 2: The layout scheme for content chunk adaptation.

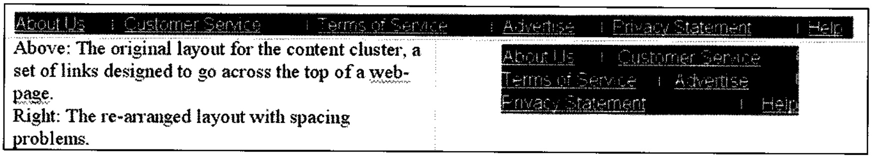


Figure 3: Example of an untidy re-arranged layout.

A new set of fifty content chunks were selected from a different set of commercial Internet sites. The chunks were rearranged using the above scheme. Ten IT proficient subjects then viewed and rated the chunks presented in both their original and re-arranged form on a browser with a screen size of 240x320 pixels. As in the previous experiment, the subjects were asked to view all the chunks before assigning a single mark between zero (unusable) and ten (excellent) for the usability of each chunk based on the same factors used in the 2000 Webby awards.

4. ALTERING WEB-PAGE METRICS

Design metrics have been categorized as relating to i) composition (e.g. word count, link count), ii) formatting (e.g. emphasized text, positioning) and iii) other general characteristics (e.g. total bytes) (Ivory *et al.*, 2000). In the next experiment, only design metrics relating to formatting were altered since to alter metrics relating to page composition requires a decision on the purpose and utility of the individual objects inside the content chunk (Paek & Smith, 1998), and altering other general characteristics such as image byte size requires knowledge of the environment the device is operating in (Mohan *et al.*, 1999), both of which are beyond the scope of this work. Based on the 30 content chunks whose layout and formatting design metrics were altered by the multimedia designer, eight easy-to-apply design metric changes were selected for investigation, below, along with the rules for their application. Some metric changes, such as the emphasis of keywords within text as suggested by the experiment in the previous section, were not implemented due to the difficulty in deciding which words should be emphasized in arbitrary body of text if such a process was to be automated.

- **Bullet list** – When there are more than three links or text objects in a list and the width of the list is less than half of the screen's width, the list is bulleted. However, lists are not bulleted if the link elements are inside table cells with a background color.

- **Justify text** – All text passages longer than 300 words are justified.
- **Center text and images** – When the width of the white space on the right hand side of a table cell is greater than a quarter of the total width of the cell, the contents of the cell are centered.
- **Expand images** – When an image is the only object in a table cell and it covers more than 80% of that cell (minus the padding), then the image is expanded, keeping the aspect ratio, to fill up the entire cell. This situation occasionally arises after the content chunk has been rearranged using the above layout scheme.
- **Add background** – When an image is the only object in a table cell and covers less than 80% of that cell, then the cell is given the same background color as the document color. Also when link elements are contained in table cells of varying widths filled with background colors different to that of the document, then the document background is set to the cell background color. Again this situation occasionally arises after the content chunk has been rearranged using the above layout scheme.
- **Omit background** – When the content chunk consists of just text on a background color or when the content chunk uses more than four colors (not including images) then the document background color is omitted.
- **Horizontal separator** – When there are more than three objects of the same type, which contain text and a single common (or no) background color, in a vertical arrangement then a separating horizontal line is inserted between the objects.
- **Vertical separator** – When the objects are arranged horizontally in a regular fashion (for example as in a table), then vertical separators are added by inserting empty columns, whose width are a quarter of that of the other columns, between the columns containing objects.

A new set of content chunks (approximately eighty) was selected from a different set of commercial Internet sites. These chunks were rearranged using the layout scheme. For each of the eight metric changes, two sets of stimuli were selected. The first set of stimuli consisted of the chunks that satisfied the requirements for the application of the metric change in question. The second set of stimuli consisted of the metric changes applied to copies of the chunks in the first set of stimuli. For each of the metric changes, the number of chunks in each set of stimuli ranged from ten to fourteen. Ten IT proficient subjects then viewed and rated the content chunks from the sixteen sets of stimuli presented in a random order on a browser with a screen size of 240x320 pixels. As in the previous experiments, the subjects were asked to view all the chunks before assigning a single mark between zero (unusable) and ten (excellent) for the usability of each chunk based on the same factors used in the 2000 Webby awards.

Table 4 : Results of applying metrics to the re-arranged content chunks. The mean and standard deviation (brackets) for the sets of stimuli are shown, along with the value of a one-sided t-test performed on the distribution of the mean marks for each of the chunks in the two sets of stimuli.

Metric Change	Bullet List	Justify	Center	Expand Image	Add B/ground	Omit B/ground	Horizontal Separator	Vertical Separator
Before Metric	5.35 (1.04)	5.42 (0.55)	6.43 (0.85)	5.85 (0.83)	5.60 (0.72)	6.55 (1.28)	5.93 (0.46)	5.35 (0.65)
After Metric	6.31 (0.54)	6.72 (0.49)	7.34 (0.64)	7.08 (0.58)	6.99 (0.81)	5.66 (1.06)	7.04 (0.83)	6.89 (0.71)
T-test value	0.01088	0.00001	0.00194	0.00024	0.00039	0.05416	0.00120	0.00002

The mean and standard deviation (in brackets) for the marks in each of the sets of stimuli are shown in Table 4, along with the results of a one-sided t-test comparing the distributions of the mean mark assigned to each chunk in the first set of stimuli against the mean mark of the same chunks after application of the metric change. From the results it can be seen that, except in the case of the omit background change, the metric changes resulted in a significant improvement (certainly in the 5% level and in most cases beyond the 1% level) in the perceived usability of the content chunks. Examination of the mark of the individual content chunks within the sets of stimuli show that application of the metric changes resulted in improved perceived usability for approximate 90% of content chunks, except in the case of the omit background change where it only improved the perceived usability of 30% of the chunk and resulted in a lower perceived usability for the other 70% of chunks.

5. CONCLUSION

The paper has introduced a simple and successful scheme for rearranging the layout of objects within a content chunk so that they are able to fit on a screen of arbitrary width. Furthermore, it was demonstrated that the perceived usability of content chunks can be improved by changing design metrics that relate to the formatting of the content. In doing so, a number of design metric changes, suitable for use in adapting PC-based web content to PDA devices, were identified along with rules for their application.

Finally, the question of which design metrics form a continuum over different devices is worthwhile of further investigation. This knowledge would allow good sets of design metric values to be predicted for new devices as they became available. From the above results and considering WAP pages, it is predicted that the number of font colors and graphics will form a continuum, both increasing with increasing screen size. However, it is suspected the use of text emphasis, centering and inclusion of horizontal separators as layout strategies will peak for device sizes similar to PDAs, with alternative strategies, such as text positioning and use of color becoming more dominant with increasing screen size.

6. REFERENCES

Banerjee, S., Gupta, A. and Basu, A. (2003), Online Transcoding of Web Pages for Mobile Devices, *Proceedings of Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2003)*, pp. 271 – 285.

Bharadvaj, H., Joshi, A. and Auephanwiriyakul, S. (1998), An Active Transcoding Proxy to Support Mobile Web Access. *17th IEEE Symposium On Reliable Distributed Systems*, October 1998.

Bickmore, T., Girgensohn, A. and Sullivan, J.W. (1999), Web Page Filtering and Re-Authoring for Mobile Users, *Computer Journal*, 42(6), pp. 534-546.

Bonhomme, S. and Roisin, C. (1996), Interactively Restructuring HTML Documents. *Computer Networks and ISDN Systems*, 28(7-11), pp. 1075-1084.

Chen, J., Zhou, B., Shi, J., Zhang, H.J. and Fengwu, Q. (2001), Function-Based Object Model Towards Website Adaptation. *Proceedings of The 10th International World Wide Web Conference*, Hong Kong.

Coninx, K., Luyten, K., Vandervelpen, C., Van den Bergh, J. and Creemers, B. (2003), Dygimes: Dynamically Generating Interfaces for Mobile Computing Devices and Embedded Systems. *Proceedings of Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2003)*, pp. 256 – 270.

Giller, V., Melcher, R., Schrammel, J., Sefelin, R. and Tscheligi, M. (2003), Usability Evaluations for Multi-device Application Development Three Example Studies. *Proceedings of Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2003)*, pp. 302 – 316.

Ivory, M.Y. and Hearst M.A. (2002), Statistical Profiles of Highly-Rated Web Sites, *CHI 2002, ACM Conference on Human Factors in Computing Systems*, CHI Letters, 2002, pp. 367-374.

Ivory, M.Y., Sinha, R.R. and Hearst M.A. (2000), Preliminary Findings on Using Quantitative Measures to Compare Favorably Ranked and Unranked Information-centric Web Pages, *Proceedings of the 6th Conference on Human Factors and the Web.*

Ivory, M.Y., Sinha, R.R. and Hearst M.A. (2001), Empirically Validated Web Page Design Metrics, *CHI 2001, ACM Conference on Human Factors in Computing Systems*, CHI Letters 3(1).

Katz, R.H. (1994), Adaption and Mobility in Wireless Information Systems, *IEEE Personal Communications*, 1(1), pp. 6-17.

Kirda, E., Kerer, C. and Matzka, G. (2001), Using XML/XSL to build adaptable database interfaces for Web site content management, XML *Software Engineering Workshop (XSE 2001), 23rd International Conference on Software Engineering*, Toronto, Ontario, Canada.

Korolev, V. and Joshi, A. (2001), An End-End Approach to Wireless Web Access, *Proceedings of the 21st IEEE International Conference, Distributed Computing Systems Workshops*, IEEE Computer Society 2001, pp. 473-478.

Ma, W.Y., Bedner, I., Chang, J., Kuchinsky, A. and Zhang, H.J. (2000), A Framework for Adaptive Content Delivery in Heterogenous Network Environments, *Proceedings of SPIE / ACM Conference on Multimedia Computing and Networking*, Feb 2000.

Mohan, R., Smith, J.R. and Li, C.S. (1999), Adapting Multimedia Internet Content for Universal Access, *IEEE Transactions on Multimedia*, pp. 104-114.

Paek, S., Smith, J.R. (1998), Detecting image purpose in World-Wide Web documents, *Proceedings IS&T / SPIE Symposium on Electronic Imaging: Science and Technology – Document Recognition*, San Jose, CA.

Ratner, J., Grose, E. and Forsythe, C. (1996), Characterization and Assessment of HTML Style Guides, *Proceedings of ACM CHI 96 Conference on Human Factors in Computing Systems*, pp. 115-116.

Schilit, B.N., Trevor, J., Hilbert, D. and Koh, T.K. (2001), m-Links: An Infrastructure for Very Small Internet Devices, *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking 2001* (Rome, Italy, July 2001), ACM Press, 2001, pp. 122-131.

Schmidt, A., Schröder, H. and Frick, O. (2000), WAP – Designing for Small User Interfaces, *ACM CHI 2000 Extended Abstracts, Conference on Human Factors in Computing Systems*, pp. 187-188.

Soumen Chakrabarti, Mukul M. Joshi and Vivek B. Tawde. (2001), Enhanced topic distillation using text, markup tags, and hyperlinks. *ACM Special Interest Group on Information Retrieval*, New Orleans.

The Webby Awards. *Official Rules + Guidelines*. http://www.webbyawards.com/main/submit/rules.html

Yang, Y. and Zhang, H.J. (2001), HTML Page Analysis Based on Visual Cues, *Sixth International Conference on Document Analysis and Recognition (ICDAR2001).*