Chapter 10

Applying the Web Interface Profiles: Empirical Examination of Web Design Guidelines

10.1 Introduction

As discussed in Chapter 5, there are a huge number of Web design guidelines available in print and on the Web. Oftentimes these recommendations are vague, contradictory, and most have not been empirically validated. This chapter presents an analysis in which metric values in the profiles are compared against established guidelines for nine Web interface aspects listed below.

- Amount of text on a page
- Length and quality of link text
- Number and type of links on a page
- Use of non-animated and animated graphical ads
- Font styles and sizes
- Unique colors and color combinations
- Download speed
- Accessibility and HTML errors
- Consistency across pages

Many of the guidelines for these aspects relate to quantitative measures that play a major role in predicting page and site quality; these measures were discussed during profile development (Section 6.5.2) and in the assessment of the example site (Section 8.5). The profiles are used to derive thresholds (i.e., ranges of acceptable metric values) for measures that are relevant to the Web design aspects above. However, thresholds for individual measures are not intended to be used in isolation; as demonstrated by the example Web site assessment in Chapter 8, the models encapsulate relationships among one or more measures that may not be reflected by individual measures. The analysis below primarily uses the good page clusters (i.e., small-page, large-page,

and formatted-page), since they provide more context than the overall page quality model. In cases where relevant measures are not included in the cluster models, the analysis uses the overall page quality model instead. Similarly, the site-level analysis (consistency across pages) uses the overall site quality model; only site-level measures are used by this model and consequently reported in the analysis. In some cases the thresholds support current design guidelines, and in other cases they counter them.

The discussion in this chapter is not intended to imply that the profiles reflect causal links. For example, one limitation of the current models and tools is that they cannot improve on poor content. However, the study in the previous chapter provided preliminary evidence that they can provide insight on how to take good content that is poorly presented and improve its presentation, thus improving users' experience in accessing that content. And, because it is possible to empirically find commonalities among the presentation elements of the highly-rated sites, this provides strong evidence that the presentational aspects of highly-rated sites that differ from those of poorly-rated sites are in fact important for good design. The assessment in this chapter examines guidelines associated with some of these presentational aspects.

10.2 Page-Level Guidelines

The following sections summarize the comparison of thresholds derived from the good page cluster and overall page quality models to Web design guidelines. Eight page-level aspects are explored, including the amount of text on a page, the number and type of links, and color usage.

10.2.1 Amount of Text on a Page (Text Element)

The literature includes the following contradictory heuristics about the ideal amount of text for a Web page. Furthermore, there is no concrete guidance on how much text is enough or too much.

1. Users prefer pages with more content over breaking content into multiple pages [Landesman and Schroeder 2000].

2. Keep text short; use 50% less text than in print publications [Nielsen 2000].

3. Break text up into smaller units on multiple pages [Flanders and Willis 1998; Nielsen 2000].

The three good page clusters (discussed in Section 6.7) provide some insight about the ideal amount of text; however, the clusters do not provide insight on whether text is broken up into multiple pages on good sites. Table 10.1 depicts key text element and formatting measures with ranges based on the cluster centroids. The ranges suggest that pages with both a small and large amount of text are acceptable; however, the profile discussion revealed that text formatting needs to be proportional to the amount of text (see Section 6.10). For example, larger pages need to contain more headings than smaller pages to facilitate scrolling. Table 10.1 shows that headings (display word count), text clustering (text cluster count), as well as the number of columns where text starts on pages (text column count) varies for the three clusters. The ranges are all significantly different across clusters as determined by ANOVAs. Thus, the ranges in Table 10.1 appear to support all of the guidelines above, although it is not possible to assess where text is broken up into multiple pages.

One related question that is often posed, especially by novice designers, is whether or not home pages should be restricted to one scroll. The overall page quality models shows that good

Measure	Good Page Cluster		
	Small-Page	Large-Page	Formatted-Page
Word Count	72.8–371.1	710.4–1008.7	218.6–517.0
Display Word Count	1.1–18.7	20.8–38.4	14.7–32.2
Text Column Count	0.6–4.4	1.9–5.7	6.4–10.1
Text Cluster Count	0.0–2.3	1.9–4.2	2.6–4.9

Table 10.1: Word count and other text element and formatting ranges for good pages. The cluster models were used to derive ranges for all of the measures.

home pages typically require from 0.69 to 3.25 vertical scrolls, which suggests that it is not the case that home pages are restricted to one scroll. However, this may depend on the content category.

10.2.2 Length and Quality of Link Text (Text Element)

Nielsen [2000] suggests that Web designers use 2–4 words in text links; however, Sawyer and Schroeder [2000] suggest that Web designers use links with 7–12 "useful" words (i.e., words that provide hints about the content on a destination page). The average link words measure for all of the good pages suggests that text links on these pages contain from two to three words. Furthermore, the average good link words measure suggests that one to three of these text link words are not stop words or the word 'click,' which suggests that they are potentially useful. Ranges for each of the three good page clusters are very similar. Hence, the data suggests that link text on good pages is consistent with Nielsen's heuristic.

10.2.3 Number and Type of Links on a Page (Link Element)

Section 5.5 provided several guidelines on the number and type of links on a page as summarized below.

1. A large number of links impedes navigation [Spool *et al.* 1999].
2. Avoid using graphical text links; they are typically ignored [Sawyer and Schroeder 2000; Spool *et al.* 1999] or may impede navigation [Scanlon and Schroeder 2000c; Spool *et al.* 1999].
3. Use corresponding text links (for graphical links) [Flanders and Willis 1998; Sano 1996].
4. Avoid within-page links, since they may be confusing [Nielsen 2000; Sawyer and Schroeder 2000; Spool *et al.* 1999].
5. Use multiple links to the same content with appropriate scent in each area [Spool *et al.* 2000].
6. Use different forms for repeated links (e.g., text, graphical text, or image) [Sawyer and Schroeder 2000].
7. Redundant links may cause confusion [Kim and Yoo 2000].

Table 10.2 provides ranges for the link, text link, redundant link, and link graphic counts on pages in the three cluster models. The table also depicts ranges for link text cluster counts and shows that the number of link text clusters are somewhat proportional to the number of links on pages. For example, pages in the formatted-page cluster appear to have the most links as well

Measure	Good Page Cluster			Overall Page
	Small-Page	Large-Page	Formatted-Page	
Link Count	12.4–41.2	35.8–64.6	54.0–82.9	
Text Link Count	5.0–28.0	24.3–47.3	35.8–58.8	
Redundant Link Count	1.4–9.1	4.6–12.3	9.0–16.7	
Link Graphic Count	5.3–14.3	5.8–14.8	12.8–21.9	
Page Link Count				0.0–5.0
Link Text Cluster Count	0.0–1.4	0.5–1.9	1.7–3.1	

Table 10.2: Link element and text formatting ranges for good pages. The cluster models were used to derive ranges for all of the measures, except page link count; the overall page quality model was used to derive the range for this measure.

as the most link text clusters. The table shows that redundant links are used on good pages and graphical links are not avoided as suggested above; it is possible that the graphical links have text on them, which is not currently detected by the metrics tool. Given the ranges for redundant links along with the ranges for text and graphical links, it appears that links are repeated in both text and image formats. However, the measures do not reveal how often redundant links correspond to those that are also graphical links.

Table 10.2 shows that good pages also contain from zero to five within-page links. The means and standard deviations for each cluster suggest that within-page links are used infrequently on pages in the small-page and formatted-page clusters; however, they appear to be used more so on pages in the large-page cluster. Examining pages with many within-page links in this cluster revealed that most were reference pages (e.g., FAQs) with numerous links to allow users to return to the top portion of pages.

10.2.4 Use of Non-Animated and Animated Graphical Ads (Graphic Element)

The following guidance has been provided on the use of graphical ads and animation.

1. Ads affect the user experience; integrate ads with content [Klee and Schroeder 2000].
2. Usability dictates that ads should be eliminated [Nielsen 2000].
3. Ads increase credibility [Kim and Fogg 1999].
4. Minimize animated graphics [Flanders and Willis 1998].
5. Avoid using animation unless it is appropriate (e.g., showing transitions over time) [Nielsen 2000].
6. Animation is irritating to users; it impedes scanning [Spool *et al.* 1999].

Table 10.3 depicts ranges for animated graphical ads in the three clusters. Pages in each cluster are likely to contain one or more graphical ads, although pages in the formatted-page cluster typically contain the most graphical ads. Good pages tend to have zero or one animated graphical ad, which suggests that no more than one graphical ad is animated on good pages. Thus, it appears that animation is used sparingly.

Measure	Good Page Cluster			Overall Page
	Small-Page	Large-Page	Formatted-Page	
Graphic Ad Count	0.1–1.5	0.2–1.6	1.6–3.0	
Animated Graphic Ad Count				0.0–1.4

Table 10.3: Graphical ad ranges for good pages. The cluster models were used to derive ranges for the graphic ad count, while the overall page quality model was used to derive the range for the animated graphic ad count.

10.2.5 Font Styles and Sizes (Text and Page Formatting)

Guidelines on the use of font styles (serif or sans serif) and ideal font sizes for text include the following.

1. Use serif fonts for faster reading by older adults [Bernard *et al.* 2001].
2. Sans serif fonts have a slight advantage over serif fonts and are more preferred [Bernard and Mills 2000; Schriver 1997].
3. Use only a few sizes from one or two typeface families; use one serif and one sans serif font for contrast [Schriver 1997].
4. Use sans serif fonts for smaller text and serif fonts for larger text [Nielsen 2000].
5. Use 14 pt fonts for older adults [Bernard *et al.* 2001].
6. Use font sizes greater than 9 pt [Flanders and Willis 1998; Schriver 1997].
7. Use 10 to 11 pt (or higher) for body text and 14 pt (or higher) for display text; use larger point sizes for serif faces [Schriver 1997].

Table 10.4 summarizes ranges for other key font style and size measures on good pages. Based on the mean for the font style measure, sans serif is the predominant font style on good pages. However, the serif font count shows that these pages may use one serif font as well. In fact, the sans serif and serif font counts suggest that one serif font face and one sans serif font face is used on good pages. These pages also use at least three font combinations (font count) – font face, size, bolding, and italics – based on the font counts. The average font size used on pages in the three clusters is 9 pt or greater, which is somewhat consistent with the guidelines above. The pages also use a slightly smaller font size (minimum font size), typically for copyright or footer text.

Contrary to recommendations for using serif fonts for display text and sans serif fonts for body text, sans serif fonts appear to be used for both on good pages. There are large correlations between body and sans serif word counts as well as medium-strength correlations between display and sans serif word counts. There are no large correlations between display and serif word counts, suggesting that serif fonts may be used for something other than headings, such as form elements. The measures do not capture whether larger font sizes are used with serif fonts and vice versa for sans serif fonts.

Measure	Good Page Cluster			Overall Page
	Small-Page	Large-Page	Formatted-Page	
Font Style				Sans Serif
Font Count	3.6–5.8	5.7–7.9	5.8–8.0	
Sans Serif Font Count				0.4–1.2
Serif Font Count				0.1–1.3
Average Font Size	10.2–11.3	10.4–11.5	9.6–10.8	
Minimum Font Size	8.8–9.1	8.8–9.1	8.8–9.0	

Table 10.4: Font style and size ranges for good pages. The cluster models were used to derive ranges for half of the measures, and the overall page quality model was used to derive ranges for the other half. The font style measure reflects the predominant font face (serif, sans serif, or undetermined) used on a page.

10.2.6 Unique Colors and Color Combinations (Text, Link, and Page Formatting)

The literature offers the following guidance on the use of colors for text, links, and other Web page elements.

1. Minimize the number of text colors [Flanders and Willis 1998].
2. Use link and visited link colors that are similar to default browser colors (e.g., shades of blue, red, and purple) [Nielsen 2000].
3. Use default browser colors for links [Spool *et al.* 1999].
4. Use no more than 6 discriminable colors [Murch 1985].
5. Use browser-safe colors [Kaiser 1998].
6. Use color combinations determined to be good (i.e., high contrast) via user studies [Murch 1985].
7. Use high contrast between background and text [Flanders and Willis 1998; Nielsen 2000].

Table 10.5 summarizes ranges for various color measures. Good pages use from one to three colors for body text (body color count). Similarly, all three of the good page clusters use from one to three colors for display text (display color count); it appears that the number of display colors is proportional to the amount of text or formatting on pages. Although it is not clear whether different colors are used for body and display text on these pages, color counts for the three clusters suggest that this may be the case. Table 10.5 shows that the total number of colors used on good pages range from five to twelve with two or three of these colors being used for links and another one to three being used for body text; there is a gap between the total number of colors and the ranges for link and body text colors suggesting that the remaining distinct colors are used for display text. It seems that good pages may use up to six different colors for text, which appears to contradict the first heuristic for minimizing the number of text colors.

Table 10.5 shows that good pages use from two to four different colors for links (link color count). Furthermore, not all of these colors are standard link colors (standard link color count); the measures do not assess whether colors similar to the standard link colors are used. These ranges suggest that good pages do not closely follow the guidance in the literature on link colors.

Measure	Good Page Cluster			Overall Page
	Small-Page	Large-Page	Formatted-Page	
Body Color Count				0.9–3.1
Display Color Count	0.7–1.6	1.3–2.2	1.7–2.6	
Link Color Count				2.3–4.3
Standard Link Color Count	0.6–1.8	0.6–1.7	0.2–1.4	
Color Count	5.4–8.1	6.5–9.2	9.7–12.5	
Browser-Safe Color Count				3.5–6.6
Good Text Color Combinations	1.3–3.5	2.2–4.5	4.6–6.9	
Neutral Text Color Combinations				0.4–2.4
Bad Text Color Combinations				0.0–1.1
Good Panel Color Combinations	0.0–0.6	0.0–0.7	0.6–1.3	
Neutral Panel Color Combinations				0.0–1.2
Bad Panel Color Combinations	0.0–1.2	0.3–1.6	1.5–2.8	

Table 10.5: Color and color combination ranges for good pages. The cluster models were used to derive ranges for half of the measures, and the overall page quality model was used to derive ranges for the other half.

Ranges for the total number of unique colors (color count) on pages in each cluster show that good pages do not adhere to the guidance of using no more than six discriminable colors. Furthermore, they do not strictly use browser-safe colors (browser-safe color count) as recommended by the literature. Good pages tend to use good text color combinations more so than neutral and bad text color combinations. However, they tend to use neutral and bad panel color combinations (i.e., thick lines or shaded areas) more so than good panel color combinations. Inspections of good pages showed that oftentimes the panel color combinations are the reverse of text color combinations. For example, the page background may be white and a blue navigation bar may be placed on top of this white background, then white text is used on top of the blue navigation bar background. Based on studies by Murch [1985], the white background with the blue navigation bar is considered a neutral panel color combination, while the blue background with white text is considered a good text color combination. This discrepancy suggests that panel color combinations may not be as important as the text color combinations on pages, since they simply represent overlapping color regions, such as a colored navigation bar placed on top of the page's background color.

10.2.7 Download Speed (Page Performance)

The time for a page to fully load is considered a critical issue for Web interfaces, and the following guidance has been offered with respect to optimizing download speed.

1. Download speed should be no more than 10 seconds [Nielsen 2000].
2. Home pages greater than 40K result in significant bailouts [Zona Research 1999].
3. Keep graphic bytes to less than 35K [Flanders and Willis 1998].

Measure	Good Page Cluster			Overall Page
	Small-Page	Large-Page	Formatted-Page	
Download Time	12.1–20.7	11.9–20.5	10.2–18.8	
HTML Bytes	4.8–14.3K	15.4–24.9K	20.6–30.1K	
Graphic Bytes				4.4–59.0K
Script Bytes	248.0–1.7K	330.0–1.8K	1.1–2.5K	

Table 10.6: Download speed and byte ranges for good pages. The cluster models were used to derive ranges for all of the measures, except graphic bytes; the overall page quality model was used to derive the range for this measure.

4. Keep the Web page and all elements under 34K [Nielsen 2000].

Table 10.6 suggests that the estimated time to download good pages is rarely under ten seconds as suggested in the literature. However, the ranges are only a rough approximation of download speed and do not take into consideration caching of elements for subsequent use in the site. The download speed for good home pages (8.0–27.5 seconds) is most relevant (assuming users enter sites through home pages) and shows that indeed the estimated download speed is rarely under ten seconds; these estimates are based on a 41.2K connection speed[1]. The estimated download speed on average home pages ranges from 8.7 to 30.1 seconds, while the estimated download speed on poor home pages ranges from 7.2 to 27.4 seconds; this difference was not significant.

The literature also suggests that home pages greater than 40K result in greater bailout [Zona Research 1999]. HTML bytes on good home pages are from 11.8K to 33.3K, which is consistent with the guidance in the literature. Table 10.6 shows that bytes on all pages are below this recommended threshold. However, ranges for graphic bytes, HTML bytes, and script bytes in the table suggest that pages do not adhere to the guidance of keeping graphic bytes and the total bytes for all page elements below 34K. Script bytes are not reported since they are negligible for this data set.

10.2.8 Accessibility and HTML Errors (Page Performance)

Analysis in Section 6.6.2 revealed that Bobby and Weblint errors are more prevalent in good pages than average and poor pages despite the guidance that Web designers adhere to accessibility principles [Clark and Dardailler 1999; Cooper 1999; Nielsen 2000; Web Accessibility Initiative 1999] and avoid making HTML errors [Bowers 1996; Kim and Fogg 1999; Fogg *et al.* 2000]. Table 10.7 shows that pages in all clusters are typically not Bobby approved and contain several accessibility errors. Both the accessibility and Weblint errors appear to be proportional to the amount of formatting on pages; it was previously reported in Section 6.6.2 that these errors are correlated with the use of tables and interactive elements. The presence of accessibility and HTML coding errors on good pages is possibility attributable to the fact that HTML inherently does not afford designing accessible and error-free pages. Another possibility is that the tools are out of date.

[1]Section 5.12.4 showed that it is rarely possible to achieve a 56.6K connection speed with the 56.6K modem, possibly due to the technical limitations of this analog modem. For 50 connection sessions to three different Internet service providers at various times of the day, the average and median connection speed was found to be 41.2K. Hence, this connection speed is used by the download speed model.

Measure	Good Page Cluster			Overall Page
	Small-Page	Large-Page	Formatted-Page	
Bobby Approved	No	No	No	
Bobby Priority 1 Errors	0.6–1.6	0.8–1.8	1.5–2.5	
Bobby Priority 2 Errors	3.1–4.3	3.2–4.4	4.1–5.3	
Bobby Priority 3 Errors				1.8–2.2
Weblint Errors	0.0–36.2	16.4–53.0	55.8–92.4	

Table 10.7: Bobby and Weblint ranges for good pages. The cluster models were used to derive ranges for all of the measures, except Bobby priority 3 errors; the overall page quality model was used to derive the range for this measure. The Bobby approved measure reflects whether a page was Bobby approved (yes or no).

10.3 Site-Level Guidelines

The following section summarizes the comparison of thresholds derived from the overall site quality model to Web design guidelines on the consistency of pages throughout a site.

10.3.1 Consistency Across Pages (Site Architecture)

The consistency of page layout and page titles in the site has been discussed extensively in the literature, and the following guidance is provided.

1. Consistent layout of graphical interfaces result in a 10–25% speedup in performance [Mahajan and Shneiderman 1997].
2. Use consistent navigational elements [Flanders and Willis 1998; Fleming 1998].
3. Use several layouts (e.g., one for each page type) for variation within the site [Sano 1996].
4. Consistent elements become invisible [Sawyer *et al.* 2000].
5. Use different page titles for each page [Nielsen 2000].

Table 10.8 summarizes several consistency aspects of good sites. Recall that the variation measures are the median coefficients of variation ($100 * \frac{\sigma}{\bar{x}}$, where σ is the standard deviation, and $\bar{x}$ is the mean) computed across the measures within each category; larger variation suggests less consistency and vice versa. The page formatting variation suggests that page layouts are very consistent (no more than 22% variation) across pages on good sites; this is consistent with the first heuristic, but contradicts the fourth heuristic. The link element and formatting variation measures suggest that navigational elements are also fairly consistent across pages on good sites; this is consistent with the second heuristic. Finally, the page title variation suggests that page titles vary considerably on good sites, which supports the last heuristic above.

10.4 Summary

This chapter demonstrated the ability to apply the profiles of highly-rated interfaces for deriving guidelines based on empirical data and for validating existing guidelines. In many cases, such as the number and type of links, graphical ads, use of animation, color usage, download speed,

Measure	Overall Site
Page Formatting Variation	0.0–22.2%
Link Element Variation	35.2–101.2%
Link Formatting Variation	0.0–43.2%
Page Title Variation	19.5–179.5%

Table 10.8: Consistency ranges for good sites. The overall site quality model was used to derive these ranges.

and accessibility, the derived guidelines contradict heuristics from the literature. However, in some cases, such as the length of link text and font styles, the derived guidelines support heuristics from the literature. This approach can be used to develop empirically-derived guidelines for other Web interface aspects discussed in Chapter 5.