How do Users Perceive Applying Web Design Guidelines?

Quentin Limbourg and Jean Vanderdonckt

Université catholique de Louvain (UCL) - Institut d'Administration et de Gestion (IAG)
Place des Doyens, 1 – B-1348 Louvain-la-Neuve (Belgium)
Phone : +32-10/47.85.25 – Fax : +32-10/47.83.24
{limbourg,vanderdonckt}@qant.ucl.ac.be

Abstract. The plethora of available Web design guidelines induces a credibility problem for designers: since most guidelines come from sources having various confidences, to what extent does applying these guidelines lead to a more usable web site that will be better accepted by users? A corpus of tested web design guidelines is therefore expected to answer that question. To reach such a corpus, five steps were performed: firstly, guidelines sources were categorized according to a source model; secondly, the most representative and interesting guidelines sources were selected according to certain criteria; thirdly, the selected guidelines were then categorized according to a general purpose model and subsequently gathered into an initial corpus; fourthly, particular guidelines subject to a user survey were identified; fifthly, these guidelines have been tested to see if users perceive any benefit in applying them. A first corpus of such guidelines was consequently composed into three parts: text-only sites, graphical sites, and framed sites. The contents of this corpus were used to develop an on-line version by a bootstrapping approach.

1 Introduction

Numerous Web design guidelines are available today throughout the channel of various sources, such as design rules, sets of guidelines, style guides, standards, and ergonomic algorithms. These guidelines are so numerous and widespread that designers wanting to use them do not exactly know how to select them and to what extent applying them could improve the site usability. The sources themselves hold different confidences: for instance, a guideline coming from a style guide does not possess the same validity than an experimentally validated guideline reported in a set.

Conducting an individual experimental study to validate each guideline would certainly provide reliable results on the effectiveness of each guideline. However, they are so numerous that this task represents a huge amount of work. Instead, we preferred to conduct a global survey where respondents have been asked to subjectively rate browsed web pages designed with and without one guideline at a time. In this way, we hoped to identify how users do perceive the application of guidelines. This research aims to develop a corpus of tested guidelines designers can thrust to improve the usability of a web site.

J. Vanderdonckt et al. (eds.), *Tools for Working with Guidelines*

In section 2, the question whether designers feel indeed a need to use a corpus of tested guidelines is raised. This question is further examined in analyzing the need for guidelines, the need for organizing them into a corpus, and the need for testing. In section 3, the process followed to reach a first version of the intended corpus is presented. This process is composed of five interdependent steps (Fig. 1): only the combination of these steps can result in a successful corpus as intended. Each step is detailed into a single section as in Fig. 1: the categorization of guidelines sources (Sect. 3), a selection of guidelines sources (Sect. 4), a categorization and gathering of guidelines (Sect. 5), a selection of guidelines (Sect. 6) and a user testing of these guidelines (Sect. 7). Finally, some conclusions are provided in a direct analysis of the strengths and weaknesses of the method (Sect. 8).

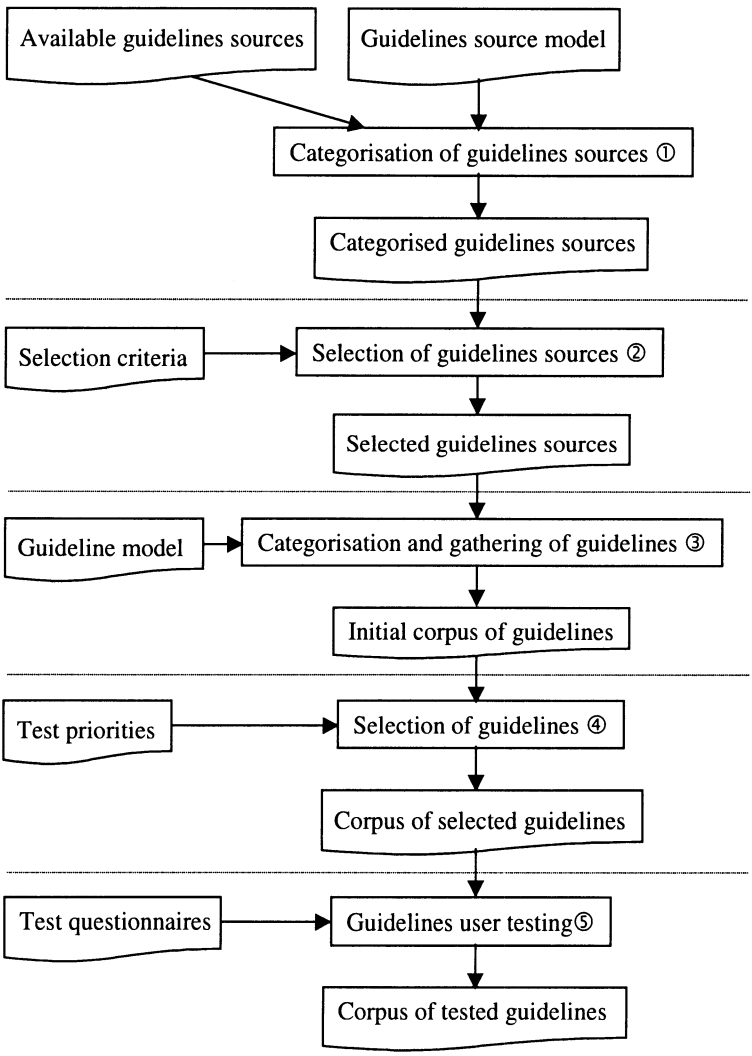

Fig. 1. The five steps of the complete process.

2 The Need for a Corpus of Validated Web Design Guidelines

2.1 The need for guidelines

Methods for Web site design do not assume to reach a unique solution. This process is not intended to be exact: as the methodology proceeds, designers rapidly face a multitude of design options involving many different aspects and allowing multiple solutions for solving a single problem. Because web site design allows alternative ways to decide each design option, there are no strict governing rules guaranteeing that one option is better than another. The absence of such rules makes the process harder for designers. However, the manipulation of guidelines can restrict the alternative palette and can shed some light on the consequences of design option [5].

Through several experiments, Borges, Morales & Rodriguez demonstrated that web site designers could effectively improve the web site usability by applying seventeen fundamental guidelines [3]. By measuring the accomplishment time for five basic tasks carried out on three sites, with and without guidelines, the authors observed some substantial time reduction in accomplishing a task (from 16% to 82%).

2.2 The need for a corpus

Guidelines can mainly be extracted from five types of sources [14]:

1. Design rules consist in a series of functional and/or operational rules governing design for a family of web pages. They are written to no longer require any designer interpretation, thus reducing the scope of alternatives to a unique one. They typically consider the physical page layout, its format, and its size.
2. Sets of ergonomic rules contain recommendations serving for a large spectrum of web sites. Each recommendation is presented with a statement, illustrated with examples, explained with some comments, and so forth. Each recommendation comes from a human consensus between the different parts using them, this consensus being theoretical, empirical or both.
3. Standards propose a sequence of functional and/or operational specifications standardizing web site design. They are promulgated by official organizations [16].
4. Style guides focus on guidelines insuring some consistency between a defined set of web pages (e.g., an organization's web site, a family of web pages for services).
5. Ergonomic algorithms try to systematize, to operationalize guidelines into tools that support designers in the web site design and evaluation process. For instance, Bobby [4] automatically checks a submitted web page for accessibility guidelines, compatibility guidelines, and for downloading guidelines.

All these types hold different properties, different confidences. Some contain general-purpose guidelines, some other focus on the Web, some both. Ratner, Grose & Forsythe compared 21 sources containing 357 unique web guidelines and 5 sources containing 270 unique general guidelines [13]. Only 53 guidelines (20%) are devoted to the web and 75% of them are surprisingly concentrated into one source. These figures show that web specific guidelines require a separate and appropriate organiza-

tion. This organization can be founded on eight empirically validated ergonomic criteria [2] since the manipulation of guidelines for traditional interactive applications could benefit from an organizing scheme based on these criteria. The Web will still demand adaptation.

2.3 The need for user testing

The most important step in building a corpus of tested guidelines consists in user testing of guidelines. A guideline is usually provided to improve some usability aspect of a web site. But since some guidelines are not experimentally validated and yet used by designers (sometimes designers select guidelines for reasons other than for usability), it is important to identify and to assess what the impact of applying these guidelines is on users. Indeed, since guidelines come from various origins (i.e. person, group, organization), their scope usually reflects why they have been introduced:

- Some guidelines are sometimes seen as the expression of a common design practice and the corpus, as a memorandum of these practices.
- Some guidelines have been introduced without proof of their effectiveness. Providing such guidelines and propagating them induce a dangerous risk: an invalidated guideline may be considered as de facto mandatory after a certain period of time because more and more designers rely on this guideline (invalid guidelines can be propagated as fast as valid guidelines can).
- Some guidelines are contradictory or inconsistent so that designers are faced to a dilemma: they are uncertain about which guideline to apply. For instance, the guideline "Try to keep default link colors" is stated on one hand as "Do not change the default colors for visited and unvisited links. An unvisited link is generally blue and a visited link is purple" and one the other hand as "Visited links should appear in red on your pages, unvisited links should appear in blue".
- Some guidelines heavily rely on the latest hardware and software technology: since both evolve dramatically, applying related guidelines is unpredictable in the future. It is probably false to believe that such guidelines could remain eternal; instead, they will vanish as soon as a more sophisticated technology replaces the existing. For instance, "Try to make the horizontal measurement of information not exceed 480 pixels" is no longer valid since recent surveys report that average screen resolution is 800x600 pixels.
- Certain designers tend to conform their design to a style guide or a standard to have a style- or standard-compliant site. But these guidelines may appear arbitrary as they are authentified by an official organization. Without valid guidelines, designers may be confused and discouraged to use guidelines.
- Tools that automatically perform syntactic analysis of the web site might provide some uncertain certification if they embody invalid guidelines.

At least for all these reasons, a corpus of guidelines needs to be tested primarily because its recent availability (more or less since 1995) is not always accompanied with empirical results bringing confidence. Any corpus of guidelines needs to be tested before it widespread false principles that could demolish the usability of web sites rather than improving it. The process followed towards a first corpus of tested web design guidelines is now presented. The sequential step presentation does not mean that steps are performed one after another: they can be partially parallelized.

3 Categorization of Guidelines Sources (①)

Guidelines sources belonging to the five types identified in section 2 are numerous and continuously growing so that providing a complete listing at a given time remains impossible. However, some studies [2,7] provide abundant references of guidelines sources that can be merged with references found in search web sites (e.g., Alta Vista, Yahoo, Lycos) or in dedicated web sites (UsableWeb). A closer examination of these sources shows that their diversity does not allow almost any type of comparison. To be able to survey the range of these guidelines sources, a domain specific categorization model is to be applied to each source. Information categorized using such a model can then be retrieved and compared. The contents of this categorization model can be interpreted as meta-information about the range of guidelines sources. The intended categorization model results from analyzing existing and foreseen attributes allowing source identification. This model, adapted from previous work, consists of the following attributes:

- Title: name of the guidelines source.
- Contents description: summary of the underlying structure, a table of contents.
- Reference: physical reference of the source (e.g., book, report, URL).
- Organization: name and type of the responsible organization (e.g., a university, a private company, a research center, a single person, an authoritative committee).
- Source type: source type as defined in section 2 (i.e., design rules, set of ergonomic rules, standards, style guides, or ergonomic algorithms).
- Publication date: date of the publication and last modification date.
- Elaborateness: approach followed by the responsible organization to build the source (e.g., survey results, free thinking, personal preference, empirical study).
- Guideline type: category to which the found guidelines belong (e.g., site specific, page contents related, navigation specific, intended for graphics, requiring Java applets, exploiting 3D animations).
- Amount of guidelines: total of guidelines found in the source.
- Guideline examples: one or many examples of guidelines found in the source.

Delving into the existing literature on human factors, guidelines, etc, a first set of 30 guidelines sources was considered as relevant because they each target specific aspects of web site design. These 30 guidelines sources were then categorized according to the above categorization model. Fig. 2 illustrates the categorization of the Yale C/AIM Web Style Guide [19].

- **Title :** Yale C/AIM Web Style Guide
- **Contents description :** style guide dedicated to web site design structured into a logical way : general philosophy, general user interface design, site design, graphics, multimedia, appendices.
- **Reference :** http://info.med.yale.edu//caim/manual/contents.html
- **Organisation :** Yale-New Haven Medical Center (academic research center)
- **Source type :** style guide
- **Publication date :** January 1997
- **Elaborateness :** elaboration based on long research (since 1991)
- **Guideline type :** site design, page design, navigation, graphics, animations, multimedia
- **Amount of guidelines :** more than 60
- **Guideline example :** *Interlaced GIF*

The conventional (non-interlaced) GIF graphic downloads one line of pixels at a time, and Web viewers like Netscape display each line of the image as it gradually builds on the screen. In interlaced GIF files the image data is stored in a format that allows Netscape (and other viewers that support interlaced GIFs) to begin to build a low-resolution version of the full-sized GIF picture on the screen while the file is still downloading. The "fuzzy-to-sharp" animated effect of interlacing is usually appealing, but the most important benefit of interlacing is that it gives the reader a quick preview of the full area of the picture. This preview effect can be misleading interlaced graphics are not faster loading than non-interlaced graphics, they just look as if they download faster because the rough preview comes up faster.

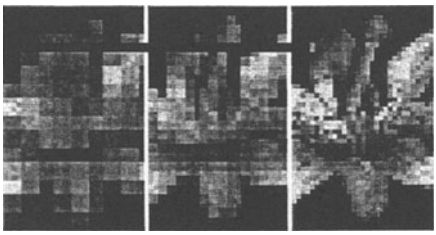

Fig. 2. Example of the categorization of a guidelines source.

4 Selection of Guidelines Sources (②)

Once guidelines sources are categorized (section 3), they can be investigated equally; their contents and their quality can be assessed according to selection criteria restricting the range of acceptable guidelines sources. These criteria are expressed in terms of the model attributes:

- Since the title is only provided as an identifier, it does not convey information relevant for selecting a source.
- The contents description does not always represent the quality of the contents, thus it is not considered.
- The reference bears little or no influence on selecting a guidelines source.

- The organization name and type are often appropriate domain indicators; we can keep organization showing experience and competence in the domain of human factors while forgetting those coming from private or personal initiatives.
- The source type is relevant as the scope and confidence of a design rule is inferior to a guideline expressed in a style guide.
- The publication date and the last modification date generally signify the attention towards the guidelines source: guidelines sources published or lastly modified before 1995 were removed.
- The elaborateness also indicates a certain confidence level: sources based on survey results, empirical studies or scientific experiments guarantee such a level.
- The guideline type often mirrors the source main goals: since we avoid hardware- and technology-oriented guidelines, sources limited to this type are left out in favor of sources related to high-level design options.
- While sources for traditional interactive applications may display a huge collection of guidelines (from 200 up to 2700), sources for web sites rarely hold more than 30 guidelines. A source including at least 50 guidelines is judged significant;
- Since copious examples facilitate interpreting guidelines, sources whitout guidelines explanation and illustration are rejected.

The above criteria need to be ranked by order of importance: the guideline type, the guidelines examples, the elaborateness, the publication/last modification date, the organization name and type, the amount of guidelines. By applying these selection criteria, the original range of 30 guidelines sources was reduced to 7 (other sources are avoided due to their poor guideline statement -second criteria- and their limited elaborateness -third criteria-):

1. Ameritech Graphical User Interface Standards and Design Guidelines [6],
2. User Interface Design for Sun Microsystem's Internal Web [9],
3. Apple Web Design Guide [1],
4. IBM Web Design Guidelines [8],
5. Yale C/AIM WWW Style Manual [19],
6. Web Site Usability: Design Guidelines [18],
7. A Jakob Nielsen's Alertbox : Guidelines for Multimedia on the Web [11].

5 Categorization, Gathering of Guidelines (③)

5.1 Categorization of guidelines

Though selected guidelines sources all satisfy the selection criteria, they do not necessarily present guidelines consistently. To make them more consistent and to facilitate their future manipulation, a previously introduced general model for categorizing guidelines is introduced. After several trials, the proposed way to transform general, informal, unstructured or differently structured guidelines into uniform, categorized guidelines was proved unsuccessful: even after having selected sources containing extensive guidelines, most attributes remained unfilled due to their poor

statement (e.g., linguistic level, ergonomic criteria, usability and utility factors, rank of importance). Filling these blank attributes by substituting ourselves to the source's author(s) is a long, hazardous, and subjective process: ignoring the context in which guidelines have been reported can falsify their expression. Even with this wish in mind, it is almost impossible to complete them uniformly in accordance with the source goals. Rather than adapting provided guidelines to the general guideline model, a specific-purpose guideline model is adapted to the sum of information available in guidelines:

- Identifier: number that univocally determines a guideline.
- Title: short sentence summarizing the main goal of a guideline.
- Statement: complete sentence describing the guideline application.
- Warrant: explanation of the guideline foundation.
- Positive examples: one or many concrete examples where the guideline is respected: these examples are accompanied with techniques and methods that effectively illustrate how the guideline is applied (e.g., HTML code, templates, files).
- Negative examples: one or many concrete examples where the guideline is transgressed.
- Bibliographic references: list of all sources where the guideline is mentioned.
- Nature: *syntactic* if the guideline can be applied, tested, verified by an automata (e.g., [4]) or *semantic* if these activities only rest on human appreciation.

- Identifier: R4.6
- Title: For graphics incorporating text, the text should be antialiased.
- Statement: Where possible, convert graphics with anti-aliasing software (for instance, Photoshop) to reduce step-like jaggedness often seen in bit-mapped graphics.
- Warrant: These "jaggies" are particularly evident and visually distracting on diagonal lines and the edges of curves and circles. Antialiase the text for better viewing.
- Positive example: 

- To see better the difference: http://www.westlake.com/gdclass/6-antialiasing/text-solid.html
- In left you will see the letter "a" without the antialiasing and in right, the antialiasing letter.
- Negative example: 

- Bibliographic reference: AMERITECH95 [6]
- Nature: semantic

Fig. 3. Example of the categorization of a guideline.

Fig. 3 depicts an example of a guideline categorized according to this model. All guidelines contained in selected guidelines sources are expressed analogously.

5.2 Gathering of guidelines

Gathering of guidelines basically consists in deciding a starting structure for the corpus and in feeding this structure with categorized guidelines.

Structuring the corpus. Although graphics are important resources in web sites, they can be inaccessible by some web browsers (e.g., Lynx). Another accessibility reason for motivating the support of text-only information is that many users with special needs are enabled to exploit text-only information by using adaptation hardware and software (e.g. auditive feedback for visually-impaired users).

A well-known accessibility problem is that formatting text into columns may confuse users: the left table of Fig. 4 is translated into a senseless meaning at the right. The related accessibility guideline here would be: "Provide alternatives to tables and frames" [16]. Tables with transparent borders are a common trick to keep a page layout appearing equally.

There is 30% chance that	Weather temperature
this event	is still increasing
will occur	tomorrow

→

There is 30% chance that weather temperature
this event is still increasing
will occur tomorrow

Fig. 4. Example of accessibility problem.

Moreover, graphics embedded in pages slow down the downloading: if this waiting time becomes too important, the user can become impatient and zap to another site. 65% of users have a major concern about the downloading time while some others prefer to get information as quick as they can so that graphics become unimportant. To support these users, a text-only version of any graphical web site is required. Most style guides also recommend this version. The corpus will therefore be divided into two parts: a part related to the textual version and a part related to a graphical version of the web site. The third part of the corpus deals with frames. Some experts estimate the usability of frames is to be disputed:

- Not all browsers support frames.
- Frames induce potential navigation problems (e.g., the "Back" button).
- When users insert a bookmark for a web page, only the home page is saved because the bookmark records the URL of all displayed frame (i.e. FRAMESET), but not the URL containing the page for data entry.
- Some browsers are unable to print information contained in a specific frame; when they are able, they are unable to print all frames in one user action;
- Frames rapidly generate scroll bars when screen display is limited.

Other experts believe that framed web sites are allowed provided experienced developers design them. Without closing the debate, we do not want to forbid frames to responsible designers. The corpus is consequently divided into three parts: guidelines for text-only sites, guidelines for graphical sites and guidelines for framed sites. This

division seems not to appear in any of the seven sources. This division is motivated by the following reasons:

- Guidelines valid for both text-only and graphical versions are duplicated (e.g., R.1.1 "At minimum, it should exist a home page" is duplicated in R.2.1).
- Developers designing a particular version can only refer to the needed part (e.g., a text-only site does not need to consider guidelines on graphical version).
- Guidelines that are unnecessary for a particular version can be omitted (e.g., "Provide Alternative Text for Images" is omitted in the text-only version).
- Guidelines inducing a potential risk for cognitive overloading are eliminated (e.g., "If you use images for navigation, make sure you also provide redundant text links" is eliminated).
- Guidelines that are common (e.g., for site design, page design, navigation) can be separated from guidelines that are specific (e.g., for graphics, for frames).

Feeding the corpus. To feed the corpus, gathered guidelines are incorporated by performing multiple operations:

- *Insertion*: if a guideline does not belong to the current pool, it is added to the relevant sections; any subsection refining the structure is created (e.g., search, user inputs); all model attributes are added whenever possible.
- *Concatenation*: if a guideline already exists in the current pool, bibliographic references are updated; each new model attribute missing in the pool is added; each existing model attribute is compared with the current one: if no difference lies between the two values, the existing value is reinforced; if a conflict separates the two values, the difference is reported; other examples are added when possible.
- *Specialization*: any guideline expressed at a very high level may be further decomposed into several specialized guidelines (e.g. by specializing a property).
- *Generalization*: any guideline expressed at low level may be partially abstracted into a general guideline (e.g., by abstracting a property).

6 Selection of Guidelines (④)

The initially set up corpus of guidelines was then submitted to a global critique. We distinguished guidelines that can be considered as admitted and guidelines that still remain open and initiate any scientific debate. It was likely that this approach will generate a certain extent of convergences and divergences, as well as shadow areas exist in the current research or in our existing scientific and statistic devices. The existence of practically no definitive rules about what should or should not be taken into account for accepting or rejecting a convergence or a divergence made the selection of guidelines a hard task. The absence of consensus on how to approach it made this task even harder, especially when considering the need for testing guidelines as pointed out in section 2. Therefore, several conventional, thus arbitrary, testing properties were enumerated to select guidelines subject to a user testing:

- Every guideline holding a contradiction should be fixed.

- Every technology-dependent guideline is to be tested.
- Every guideline having warrant attribute unspecified should be tested;
- Every guideline referring to a well established theory will not be tested.
- Every guideline based on a proved principle or derived from such a principle will not be tested,
- Every guideline hard or subjective to be tested will be discarded.

Fig. 5a shows a guideline based on a user preference for fast page access. Having this in mind, two other guidelines attaching a file type to optimal compression techniques (Fig. 5b and Fig. 5c) might be tested by derivation. All guidelines contained in the initial corpus were then marked as subject to testing. Marked guidelines were then selected for user testing in the next step.

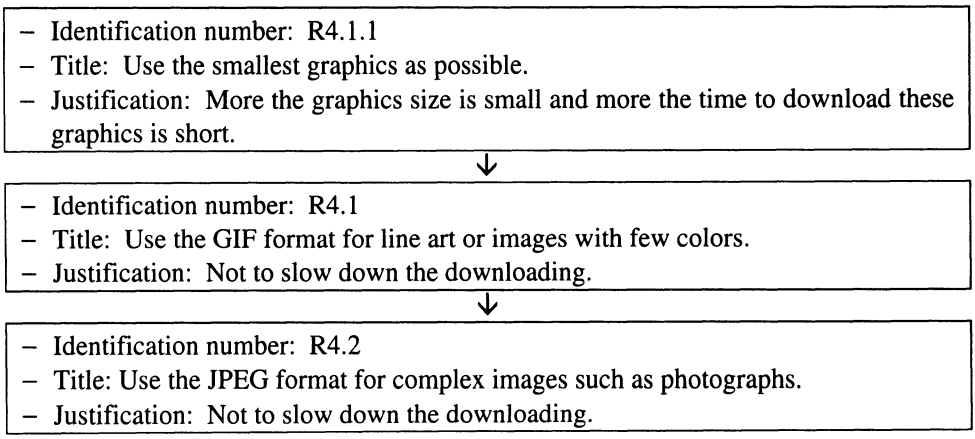

Fig. 5. Example of guidelines derivation.

7 Guidelines User Testing(⑤)

Guidelines selected for testing were finally submitted to a user testing method. The results of applying this method are presented and discussed hereafter. Since the main goal of this paper is to present the approach for a corpus of tested web design guidelines, the main characteristics of the survey are only briefly outlined.

7.1 Method

A testing method is said theoretical when the guideline is motivated by theoretical scientific evidence like cognitive psychology, theory of human factors, software ergonomics. Since sometimes various theoretical reasons can compete one with another and since real users' preferences are not always compatible with theoretical deduction, a testing method can be considered empirical when the guideline is approved on basis of acquired figures (e.g., experimental measures). In this kind of method, usability questionnaires that report experiences on measurable variables are frequently used. Some of them are of general purpose such as Software Usability Measurement

Inventory (SUMI) [15] or Questionnaire for User Interface Satisfaction (QUIS) [12] while others are specific for Web sites such as Web site Analysis and Measurement Inventory (WAMMI) [17]. These questionnaires were intended to test the usability of a user interface for a traditional interactive application or for a web site. To test the corpus of selected guidelines, testing questionnaires were chosen instead of usability questionnaires:

1. Usability questionnaires are more intended to test whether a particular user interface is usable or not. To reach a guideline, this test needs to be reproduced with other user interfaces, under varying circumstances, thus requiring intensive web site re-development and testing.
2. Testing questionnaires are more intended to directly let users estimate whether a guideline is valid or not by appreciating pages where the guidelines is applied or not or with variations. In this sense, they are closer to our testing goal.
3. The empirically measurable quality criteria proposed by Shneiderman [12], i.e. time to learn, speed of performance, rate of errors by users, subjective satisfaction, retention over time, time to learn and speed of performance are often used to validate guidelines [3,5]. The subjective user satisfaction has been chosen because it is the easiest criteria to measure and it does not subsume any software.
4. This other viewing angle can confirm or infirm guidelines that have been validated by other criteria. Performance does not necessarily match user performance.

Basically, seven questionnaires were developed and submitted to users:

- Questionnaire #1 collects demographic data (e.g., sex, age, country, background, experience level) and optional personal information (e.g., e-mail).
- Questionnaire #2 tests the corpus structure by asking relevant question.
- Questionnaire #3 asks questions about the page components and links to graphics.
- Questionnaire #4 asks questions about the page components, the interaction objects and graphical links (i.e. graphical buttons).
- Questionnaire #5 asks questions about conveying information through graphics, graphical links (i.e. image maps) and navigational aid.
- Questionnaires #6 and #7 ask questions about global and local navigation, page size and layout, page format for printing/saving purposes.

Fig. 6 reproduces some typical questions raised in these questionnaires. Each question is identified by a three-digit number: the first digit corresponds to the questionnaire number, the second digit is related to a questionnaire section, the third digit locates the question within this section. For instance, 423 stands for question 3 in section 2 related to page components in questionnaire #4 about graphical web sites. For reference, the full electronic questionnaire is available on line at http://www.info.fundp.ac.be/~vesale/PVWG/Quest. These questions are formulated so that the following requirement are met: questions should be clear, questions should not be directive, questions should be neutral, questions should not be embarrassing, questions should not be hypothetical, questions should minimize the prestige bias, the learning effort. Moreover, some questions hold some subjective, open and qualitative character. This seems appropriate since the validation questionnaire is aimed to collect us-

ers' opinion (hence, subjective) that is representative (hence, open) about guideline (hence, qualitative).

- Questions 313-413-513-613-713: What do you think about a web site accessible in both graphical and text-only versions? The idea behind this question is to identify what the user satisfaction is for having various web site versions.
- Questions 314-414-514-614-714: What do you think about a web site having two different versions at the same time: One with "frames" and another without "frames"? The idea behind this question is to identify what the user satisfaction is for having framed and unframed versions.
- Questions 323-423: What do you think about having a link to a home page?
- The idea behind this question is to try to know the average importance of a link to a home page on a scale with 7 levels (0=unuseful, 3=practical, 6=essential)
- Question 551: without using mouse (unless for scrolling), how many clickable regions do you see in the following image maps?
  Image map 1: ...
  Image map 2: ...
  Image map 3: ...
  Image map 4: ...
  The idea behind this question is to submit to the user 4 rather differently designed image maps and to discover the recognition criteria.
- Question 552: In the above image maps, where are the clickable regions easily perceived and why?

Fig. 6. Example of questions raised in the questionnaire.

7.2 Results and discussion

Survey participants were solicited as follows: announcements were sent on newsgroup, to different mailing lists, to some organization (e.g., members of the Computer Science Department of Loughborough University of Technology, UK, and students of Institut d'Informatique, Belgium) and by paper mail (all members of FUNDP). 164 people responded to the survey over a one-month period from May 15, 1998, to June 30, 1998. Like most surveys on the Internet, the sampling suffered from non-random selection and self-selection. Respondents responded via the electronic questionnaires on the Internet, through e-mail for those who subscribed to the mailing lists and postal mail for people who do not use Web or e-mail frequently.

Most respondents came from Europe and USA (USA+Canada=36%, Europe-Belgium=31%, Belgium=27%, Other=6%). Their average age was 33 years old and their professional background was mostly related to Computer Science (82%) versus non-Computer Science (18%). Their computer experience level is relatively high (μ=4.8 on a 0-6 scale) as well as their Internet proficiency (μ=4.6 on a 0-6 scale). The responses provided by participants to questionnaires were examined in order to deduce and motivate a testing score for each concerned guideline as follows:

$$\text{score} = \begin{cases} 0 \text{ if nothing can be deduced from user responses} \\ 1 \text{ if some elements came in favour of the guideline} \\ 2 \text{ if many elements came in favour of the guideline} \\ 3 \text{ if almost all responses concur in favour of the guideline} \end{cases}$$

Each guideline is examined across responses from the questions and each question provided responses to examine several guidelines. Here are four examples.

Testing the corpus structure. Through questions 313, 414, 513, 613, 713, 7% of respondents judged important that a same web site enables users to choose between a text-only version and a graphical version; 41% of respondents evoked excessive downloading times and 17% reported several accessibility troubles concerning the graphical version. Through questions 314, 414, 514, 614, 714, 82% of respondents found useful a choice between a graphical version and a framed version; 28% of participants are really concerned about frames (see reported problems above) and suck about them. Therefore, a large portion of users seems to adhere to the suggested repartition into three parts (testing score = 3): a text-only version is always required, a graphical version is not forbidden, but more discussion should be initiated on frames.

Testing a text-only guideline. This test concerns guideline R.2.1.2.1 entitled "The WWW page should include the authors' name. Through questions 323 and 423, the utility (μ=0,413) of providing the site author's name is not considered as fundamental. The presence of the author's name is consequently considered as practical, nothing more (testing score=1).

Testing a graphical guideline. This test concerns guideline R.3.1.2.3 entitled "For image maps, clearly delineate the clickable region". In question 551, 4 images maps were submitted to users' appreciation:

1. Image map #1 represented a country map with 12 red region names and 2 blue labels: 32% of respondents identified 12 clickable regions, 52% identified from 10 to 15 clickable regions, while 32% did not perceive any region;
2. Image map #2 represented a world map: 52% did not perceive any region where 36% identified between 5 and 7 regions (probably, the continents);
3. Image map #3 showed a picture of the Rushmore Mounts where presidents' heads are sculpted with their names: 80% did recognize the clickable regions while 12% did not perceive anything;
4. Image map #14 depicted a navigation bar equipped with 13 navigation buttons and a logo: 92% were able to discern 13 or 14 clickable regions, 84% identified the buttons, while only 4% did not perceive anything.

Through question 552, 72% of respondents quoted borders as a clear way to recognize a clickable region (in image map 4), 28% spoke about a label (in image map 3) and 28% identified the color code and the format (in image map 1). Through question 553, 64% of respondents believed borders are the best way to delineate a clickable region and 20% prefer color and style combination to delineate it. The clickable regions are consequently vital to delineate (testing score=3), for example with borders, color, style by order of preference.

Testing a frame-based guideline. This testing concerns guideline R.1.1 entitled "Use frames if you really need them". Through questions 314, 414, 514, 614, 714, 29% of respondents reported that frames still cause a real web accessibility hindrance while 12% believed frames improve page layout provided they are properly implemented. Since one third of visitors are allergic to frames, their usage was only partially recommended (testing score=2). By repeating this process on selected guidelines, they have been attached to testing scores as indicated in Table 1.

Table 1. Results by validation score

Testing score	Amount of guidelines
3	38
2	26
1	8
0	8

8 Conclusion

Throughout the five steps discussed in sections 3 to 7, 64 guidelines out of 80 have been tested acceptable to users (testing score=2 or 3), where 16 guidelines were tested unacceptable (testing score=0 or 1). Testing guidelines with such scores increases the certainty and the confidence in applying these guidelines; after been applied, they can be certified with a minimum expected user satisfaction level. This level does not necessarily meet results and conclusions provided by measuring user performance. It is therefore important to compare these two series of results to highlight differences that might arise between preference and performance.

It is important to note here that the main assumption made in this user survey was that a user is able to express his/her own acceptance of a guideline by interacting with pages presenting the user with various applications of a guideline and by expressing the user subjective satisfaction for each alternative. Someone might express that a guideline applied partially or totally in these pages is not acceptable but this does not necessarily imply that the guidelines would not have been of practical benefit during the actual development of a web site.

In this sense, a guideline is said acceptable only and only if a test score of 2 or 3 is obtained, which means that, the average user subjective satisfaction is acceptable. This will certainly not replace extensive experimental results on guidelines. But in on the other hand, the process reached a first corpus of guidelines which are consistent and uniformly described.

As mentioned before, ergonomic algorithms embodied in software tools should now consider these testing scores to evaluate a web site across a checklist of guidelines: settings should be allowed to specify the scope of evaluation (preference versus performance). A software tool cannot unfortunately encompass all guidelines tested in this approach since some of them are of semantic nature. Only syntactic guidelines can be mechanized. This is why it seems important to us

- To confirm/infirm guidelines that have been accepted or not accepted in this survey with other studies.
- To integrate the current corpus into any set of newly obtained guidelines to allow expanding.
- To open the structure and the contents to any advance in research and development.

Anyway, the approach here described lead us to consider a first sketch of corpus containing web design guidelines whose structure and contents were partially accepted. This corpus has been itself used to check the electronic version of the corpus both in French and in English, thus allowing a bootstrapping approach. This web site is available at http://www.info.fundp.ac.be/~vesale/PVWG/Guide.

Acknowledgments

The authors would like to thank Pierre Badot and Valery Detez for conducting the users survey, Pierre Badot and Valéry Detez for gathering and organizing guidelines, and Prof. James Alty, Loughborough University of Technology for his advice and all respondents for taking the time to respond to the survey and for providing insightful comments and appreciation. The authors would like to thank the anonymous reviewers as well.

References

1. Apple Web Design Guide. Apple Computers Inc. (1997)
2. Bastien, J.M.C., Scapin, D.L.: Evaluating a User Interface with Ergonomic Criteria. Int. J. of Human-Computer Interaction 7 (1995) 105–121
3. Borges, J.A., Morales, I., Rodríguez, N.J.: Guidelines for Designing Usable World Wide Web Pages. In Companion Proc. of CHI'96. ACM Press, New York (1996) 277–278
4. Bobby Version 2.0, HTML Web analyzer. Center for Applied Special Technology (CAST), National Universal Design Laboratory Peabody (1997). Accessible at http://www.cast.org/bobby
5. Comber, T.: Building Usable Web Pages: An HCI Perspective. In Proc. of AusWeb'95. Norsearch Ltd Ballina (1995) 119–124
6. Detweiler, M.C., Omanson, R.C.: Ameritech Web Page User Interface Standards and Design Guidelines. Ameritech Corp. (1996)
7. Grose, E.M., Forsythe, Ch., Ratner, J. (eds.): Human Factors and Web Development. Lawrence Erlbaum Associates, Mahwah (1998)
8. IBM Web Design Guidelines. International Businness Machines (1997)
9. Levine, R.: User Interface Design for Sun Microsystem's Internal Web. Sun Microsystems Inc. (1996). Accessible at http://www.sun.com/styleguide
10. Lowney, G.: Design Guidelines to Make Sites Accessible. In Proc. of HFWeb3 (1996).
11. Nielsen, J.: Guidelines for Multimedia on the Web, A Jakob Nielsen's Alertbox (1995). Accessible at http://www.useit.com/alertbox/9512.html
12. Questionnaire for User Interface Satisfaction (QUIS). Human-Computer Interaction Laboratory, University of Maryland (1996)

13. Ratner, J., Grose, E.M., Forsythe, Ch.: Characterization and Assessment of HTML Style Guides. In Companion Proc. of CHI'96. ACM Press, New York (1996).
14. Scapin, D.L.: Organizing Human Factors Knowledge for the Evaluation and Design of User Inter-faces. Int. J. of Man-Machine Studies 2-3 (1990) 203–229
15. Software Usability Measurement Inventory (SUMI). Human Factors Research Group, University of Cork (1997).
16. WAI Accessibility Guidelines, Web Accessibility Initiative. World Wide Web Consortium, Geneva (1998). Accessible at http://www.w3.org/wai
17. Website Analysis and MeasureMent Inventory (WAMMI). Nomos Management AG (1997). Accessible at http://www.nomos.se/wammi/index.html
18. Web Site Usability: Design Guidelines (1997). Accessible at http://infoweb.magi.com/~asd/guidelines/guide.htm
19. Yale C/AIM Web Style Guide. Yale-New Haven Medical Center (1997). Accessible at http://info.med.yale.edu/caim/manual/contents.html.