**DOCTOR OF SCIENCES**

# Methodology for automating web usability and accessibility evaluation by guideline

Beirekdar, Abdo

*Award date:*
2004

*Awarding institution:*
University of Namur


**Facultés Universitaires**
**Notre-Dame de la Paix**

# A Methodology for Automating Guideline Review of Web Sites

**Abdo Beirekdar**

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Sciences
(Computer Science Option)

- August 30th, 2004 -

| | |
|---|---|
| Director: | Professor M. Noirhomme-Fraiture |
| Co-director: | Professor J. Vanderdonckt, Université Catholique de Louvain, Belgium |
| Jury: | Professor F. Bodart |
| | Professor J.-L. Hainaut (President) |
| | Professor Ch. Kolski, Université de Valenciennes, France |
| | Professor Ph. Palanque, Université Paul Sabatier - Toulouse III, France |

**Institut d'Informatique**
NAMUR

# Chapter 1
# Introduction

## 1.1 The Situation

Due to the proliferation of Internet and the explosion of its user population, web sites are becoming a major source of information, communication, and collaboration in the modern society. As the use of Web sites is widening, and the number of users who approach them is increasing continuously, it is important that information be easily reachable by all. More precisely, the user population is expanding in age (ranging from young users to elderly people), in expectations (ranging from private use for leisure to professional use), in information needs (ranging from simple information to compound multimedia resources), in task types (ranging from basic text searches to complex problem-solving methods), and in user abilities (ranging from the able-bodied person to any person with special needs, such as for motor - auditory - or visually - impaired persons). The difficulties in providing such a universal access can be addressed through the application of the principles of Web ergonomics. In fact, the most considered aspects by researchers when speaking about ergonomics are usability and accessibility (U&A). For this reason, we will explicitly discuss U&A in the remaining of this dissertation.

U&A are two concepts that can be referred not only to Web sites, but also to any general User Interface (UI). Usually, we can make a distinction between WIMP (Windows, Icons, Pointer, and Mouse) interfaces and Web interfaces, in part because the nature of these interfaces differ and in part because the usability methods developed have often only been applied to one type or the other in the literature. This work deals with U&A of Web interfaces only. More precisely, it deals with automating Web U&A evaluation. The main reasons for this choice are:

- U&A of Web interfaces are less explored that U&A of WIMP interfaces [Ivory 2001].
- The data needed by automated usability methods can be obtained more easily from Web sites because most resources are available publicly via the site or on the Web server (Html code, Log files, etc.), whereas WIMP applications are generally locked as binary executable files.
- Until now, the majority of Web sites are built using HTML language, which augments the possibility of applying a given method on a large scale; whereas for WIMP applications, they are built with a multiplicity of different, and generally incompatible, languages (Java, Pascal, VB, etc.).

A *Website* is an interactive software system composed of Web pages that can be seen as dialogue windows in traditional WIMP systems. It interacts with at least two different kinds of users: end users trying to achieve some goal and

developers/maintainers striving to keep the system working and improving it [Brajnik 2000].

End users can be characterized in terms of:

- Goals and tasks: e.g. information seeking, choosing where to buy some specific product, buying it, writing a book review, etc.
- User characteristics: user behavior during information seeking processes is strongly affected by users (culture, language, previous knowledge of the field, experience in using the Web, etc.).
- Technology: end users interact with the Web site through a layer of technology that is not under control by the Web designer, such as browsers, protocols, operating system platforms, interaction devices (screens, speaking devices, reduced telephone keyboards, etc.), network connections, etc.

As for developers and maintainers, amongst their activities, a prominent role is played by actions that include: corrective maintenance (i.e. fixing problems with the Web site behavior or inserting missing contents), adaptive maintenance (i.e. upgrading the site with respect to new technologies), quality maintenance (i.e. improving the site behavior or content), and preventive maintenance (i.e. fixing problems in behavior or content before they affect users).

In fact, information seeking through browsing is a process that almost all Web sites must support. Unfortunately, it is also a difficult task to model and support because it encompasses complex cognitive, social and cultural processes [Allen 1996] spanning through interpretation of textual, visual, audio messages, selection of relevant information and learning.

As this work focuses on U&A of Web sites, we discuss these two concepts in the following sections.

### 1.1.1 Usability

Usability of a Website is determined by user satisfaction, ease of learning and remembering its organization and functionalities, user effectiveness, efficiency and likelihood of errors while performing the tasks the site has been designed for, like finding information needed, or completing e-commerce operation. A definition of usability by the standard ISO 9241 [1999] is: "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments", where:

- **Effectiveness** means "the accuracy and completeness with which specified users can achieve specified goals in particular environments;
- **Efficiency** means "the resources expended in relation to the accuracy and completeness of goals achieved";
- **Satisfaction** means "the comfort and acceptability of the system to its users and other people affected by its use".

The most important properties for Web site usability (most of them taken from [Fleming 1998]), that are often taken into consideration, are:

- Consistency of presentation and controls across the site (figure 1.1), natural organization of information (clear structure, systematic labels, clear and meaningful labels);
- Contextual navigation, in terms of environment, "type" of users, particular devices, etc; someone often considers also how much information is given for providing a context to the user (where is he, where he can go, and so on);
- Robustness, i.e. how well the Website handles technology used by users that has not been foreseen by developers;
- Flexibility (e.g. availability of graphic and textual versions, redundant indexes and site maps, duplicated image map links);
- Functionality (i.e. support of user goals) etc.

These properties need to be decomposed into more detailed ones that can be assessed in a simpler and perhaps more standard way in order to make them operational. They can be considered in the Web site evaluation, in order to determine whether it is usable or not.

Usability is not the only important aspect of the site: obviously its content/functionality, and its popularity contribute even more to the success of a site. However, if users can do the same thing with two different sites, they will choose the one that is more effective, efficient and satisfactory [Leporini 2003].

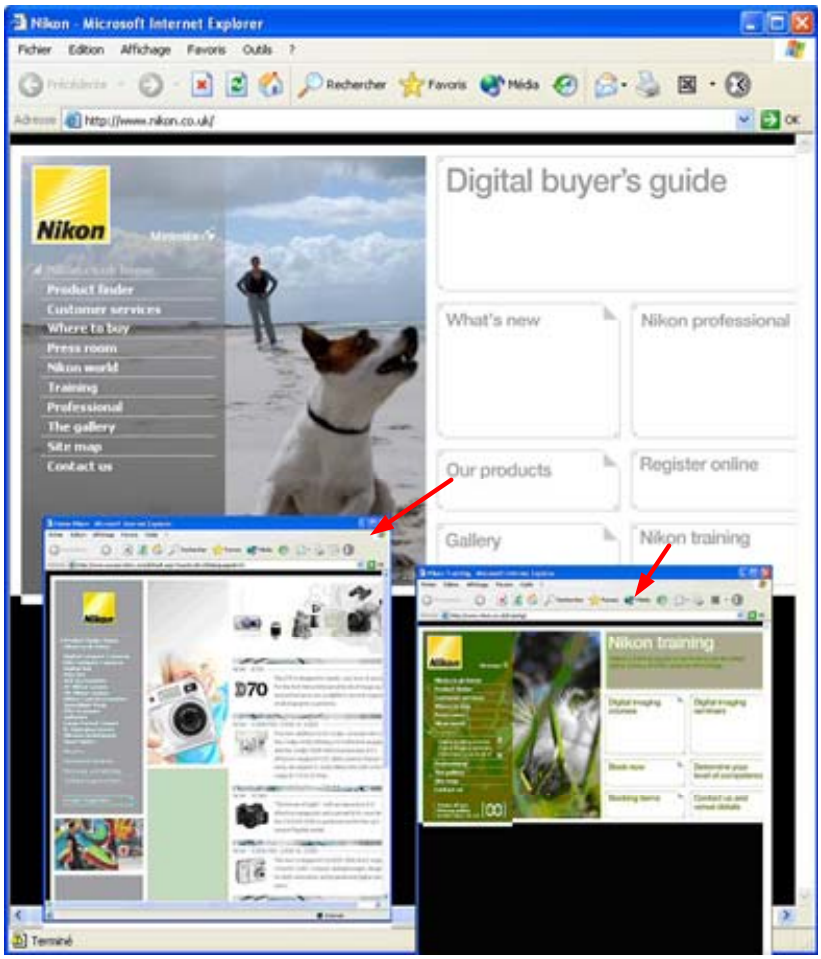


**Figure 1.1**: www.Nikon.co.uk. Navigation is inconsistent throughout this site. Browse for a few minutes and you will be transported through a variety of different divisions and, unfortunately, different site designs and incompatible navigation mechanisms.

### 1.1.2 Accessibility

In order to refer to accessibility aspects, we have to consider that many users may be operating in contexts of use that are very different from those of general use: they may not be able to see, hear, move, or may not be able to process some types of information easily or at all; they may have difficulty in reading or comprehending text; they may not have or be able to use a keyboard or mouse; they may have a text-only screen, a small screen, or a slow internet connection; they may not speak or understand fluently the language in which the document is written; they may be in a situation where their eyes, ears, or hands are busy or interfered with (e.g., driving to work, working in a loud environment, etc.); they may have an early version of a browser, an entirely different browser, a voice browser, or a different operating system. The following figures illustrate the influence of operating context on information accessibility by a blind person.

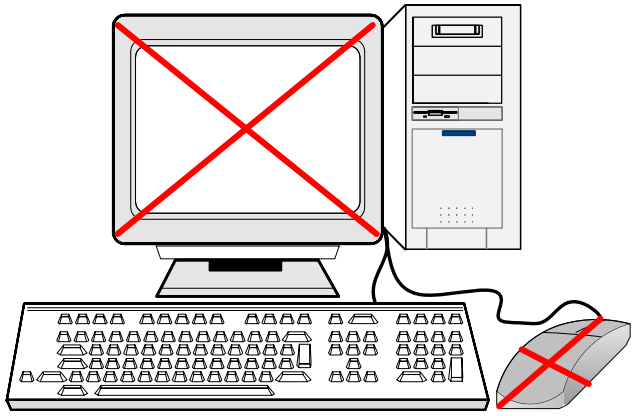

**Figure 1.2a**: How a blind person perceives a typical PC. Without screen he does not have global vision of the Web page. Without mouse, his navigation ability is limited to tabulation.

**Figure 1.2b**: Typical PC configuration for a blind person. It is composed of a vocal synthesizer, refreshable Braille device[1] and a keyboard (ASCII, Braille or specialized).

In the context of Web site design, accessibility is a measure of how easy it is to access, read, and understand the content of a Web site. A Web site can be said to

[1] Devices with small pins that can be raised or lowered to form Braille characters which the deaf-blind individual can feel.

be accessible if it can be used from everyone, regardless of disability [Brewer 2003].

Accessibility is complicated by the fact that a Web site can be interpreted in different ways by different browsers (Figure 1.3) and on different platforms. For example, in addition to people with disabilities who explore Web pages by using screen readers (with voice synthesizer or Braille display), we can include those using low-bandwidth technology like cellular phones, black and white screens, speaking browsers via telephone, etc.

In order to obtain more accessible Web sites, the Web designers should follow some simple criteria and guidelines, which do not limit the graphical features, such as images and icons, as a way to improve accessibility, but rather provide text equivalent information combined to each multimedia content.

Most accessibility issues are taken into account especially by W3C (World Wide Web Consortium) in the project Web Accessibility Initiative (WAI). In fact, Tim Berners-Lee (W3C Director and inventor of the World Wide Web) has defined accessibility as "The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect".

In the Web Guideline Initiative (WAI) activities, a set of specific guidelines and recommendations has been defined: "Web Content Accessibility Guidelines 1.0" [WAI 1999]. These guidelines explain how to make Web content accessible to people with disabilities. The guidelines are intended for all Web content developers (page authors and site designers) and for developers of authoring tools. The primary goal of these guidelines is to promote accessibility. However, following them by designers will also make Web content more available to all users, whatever user agent[2] they are using.

### 1.1.3 Relationship between Accessibility and Usability

Webster defines accessibility as "capable of being reached" and it defines usability as "capable of being used". Therefore, persons with disabilities must be able to reach Web sites, yet everyone (disabled and non-disabled) must be able to use them. To illustrate, a website may be accessible to visitors with vision impairments by having text descriptions on images so screen readers can decipher them. However, if the image does not represent anything on the site, then it is useless to all visitors. Similarly, if persons with dexterity impairments can navigate between links using TABS, then accessibility exists. But if links do not lead to anywhere, usability does not exist.

In fact, accessibility and usability are closely related, as they both improve satisfaction, effectiveness, and efficiency of the generic user population. They are closely related, accessibility is aimed more specifically at making a Web site

[2] Software to access Web content, including desktop graphical browsers, text browsers, voice browsers, mobile phones, multimedia players, plug-ins, and some software assistive technologies used in conjunction with browsers such as screen readers, screen magnifiers, and voice recognition software.

simply more available to a wider user population, while the goal of usability is to make users' experience with the Web site more efficient and satisfying.

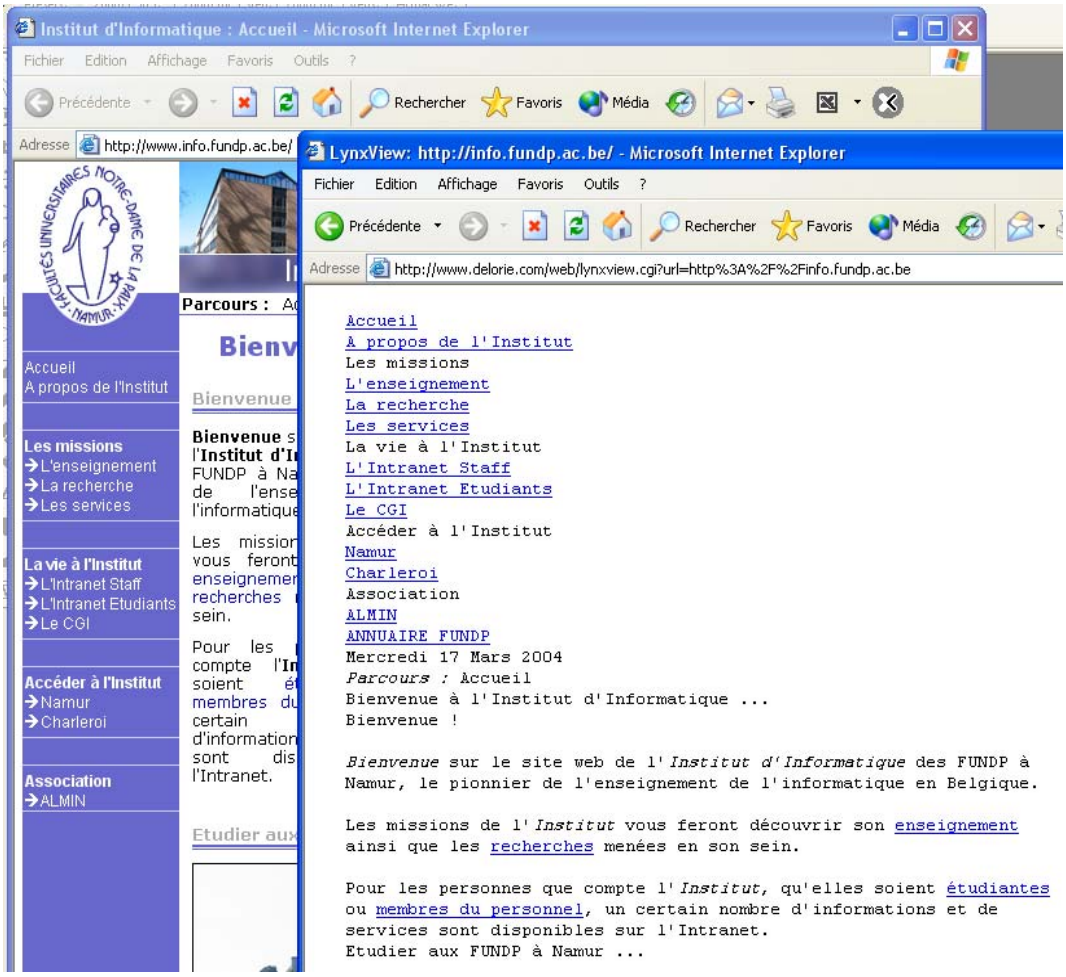


**Figure 1.3**: presentation of a Web site by IE6 (back) and by the text-mode Web browser Lynx (front). Lynx browser augments information accessibility but information layout (thus usability) is seriously affected.

In other words, accessibility issues are sometimes considered to be a matter of accessing information, while usability issues to be a question of facility in information navigation. For instance, if graphical links do not have ALT content, it is an accessibility question. If graphical links have "poor" ALT contents (e.g., "click here"), or all the contents are equal among them (e.g., "go to paragraph"), it is a usability issue.

Koch [2004] examined the U&A relationship and underlined these two principles:

- Ideally, any site should be accessible. Any device should be able to access the content and navigation of any site.
- Ideally, any site should be usable. Though basic usability is ascertained by clear texts, unambiguous forms and a simple navigation, adding some JavaScript and CSS can dramatically improve a site.

Usability interplays with accessibility in two ways. First, assistive applications, such as screen readers and TAB navigation, make Web sites more usable for persons with disabilities. Obviously, without a screen reader, blind users could not learn to navigate a Web site quickly and certainly wouldn't enjoy visiting it.

Second, assistive applications themselves have to be usable. If skipped links result when pressing TAB, then the user cannot be quick or enjoy the action and thus the Web site. As we can see, accessibility and usability are a necessary duel for persons with disabilities to fully benefit from the Web.

As for priority, a Web site should be first accessible before being usable; otherwise, usability does not matter because we can not judge it until we can access the Web site. Achieving universal accessibility is part way towards the goal of achieving universal usability. However, an "accessible" Web site may still have serious usability problems that make it equally difficult for any person, disabled or non-disabled, to use the site for its intended purpose. Usability focuses on making software applications and services easy for all people to use, so good usability practices are also good accessibility practices.

## 1.2 The Problem

The Web is intended to enable broader access to information and services than was previously available. Therefore, Web site U&A are very important issues. Unfortunately, U&A continue to be pressing problems. An estimated 90% of sites provide inadequate usability [Forrester Research 1999], and an estimated 66% of sites are inaccessible to users with disabilities [Jackson-Sanborn et al. 2002]. Although numerous assistive devices, such as screen readers and special keyboards, facilitate Web sites accessibility, these devices may not improve a user's ability to find information, purchase products and complete other tasks on sites. For example, sites may not have links to help blind users skip over navigation bars, or sites may not enable users to increase the font size of text, so that they can read it.

The USA's National Organization on Disability estimates that 35% of Americans have a minor disability and 25% has a severe disability [TecAccess 2003]. Moreover, since many populations like Europeans and Americans are getting older, they face diminishing vision and hearing, arthritis, and osteoporosis. Although not all of these people are affected by Web accessibility, a great many are prevented from receiving information in a format suited to their needs.

Via the Internet, people with disabilities now have easier access to books and information and they are able to shop unassisted. They are in a position to improve the quality of their lives if the Web sites to do so are made accessible to/usable by them.

### 1.2.1 Reasons

This lack of U&A is due to many reasons:

- In many cases, site designers have little or no experience in ergonomic web design; all their attention is focused on publishing their sites, generally using web design tools that do not integrate any kind of assistance concerning designing usable sites;
- The compatibility problem between different web technologies (editors, browsers, platforms, etc.) makes it difficult and time and effort consuming to design sites compatible with all technologies;

- Web browsers do not immediately support new publishing technologies, and obviously, old browsers that are still used by many users do not support them.
- Many sites change their content continuously by adding, removing or updating web pages. These changes could be done by different persons and are generally a main source of usability problems, especially for very-large web sites.

### 1.2.2 Solution1: User testing

One way to ensure the U&A of Web sites is via users testing. Nielsen [1998] claims that it takes 39 hours to test a Web site usability for the first time, including planning the test, defining the test tasks, recruiting test participants, conducting a test with five participants, analyzing the results, and writing the report; with experience, this time can be reduced to 16 hours. Nielsen further claims that a usability test with five participants will typically reveal 80% of the site-level usability problems (e.g., home page, information architecture, navigation and search, linking strategy, etc.) and 50% of the page-level problems (e.g., understandability of headings, links, and graphics). The author advocates increasing page-level usability through other methods such as heuristic evaluation. Contrary to these findings, Spool and Schroeder [2001] have shown that five participants only find 35% of usability problems when the participants do not complete the same tasks. This situation is exacerbated when the size of the Web site is very large, thus, it appears that testing may not be a viable method for accelerating and improving the Web design process.

### 1.2.3 Solution2: U&A Guidelines

As a complement to testing, many detailed U&A guidelines have been developed for both general user interfaces [Open Software Foundation 1991; Smith and Mosier 1986] and for web design [WAI 1999; Comber 1995; Lynch and Horton 1999]. A guideline is a principle of conception and/or evaluation to be observed in order to have and/or to insure an ergonomic human-machine interface [Vanderdonckt 1994].

The application of such guidelines has already been proved to have a positive impact on U&A [Comber 1995]. For instance, Borges et al. showed that the average time to carry out 5 tasks on a web site respecting 17 guidelines has been reduced by at least 16% [Borges et al. 1996].

However, this does not necessarily imply that a Web site that does not consider guidelines is unusable. Nor it is proved that a web site addressing all guidelines is the most usable and accessible site. In other words, guidelines are necessary, but insufficient. They should not be considered in isolation: often guidelines need to be supplemented by a suitable method and a clear process that leads them to unambiguous interpretation. In addition, designers have historically experienced difficulties following design guidelines [Borges et al. 1996; Lowgren and Nordqvist 1992; Smith 1986; Chevalier and Ivory 2003; de Souza and Bevan 1990; Scapin et al. 2000]. This is due to reasons extensively discussed in [Vanderdonckt 1999] such as, but not limited to:

- Guidelines are expressed as general recommendations independent from context. Their application consequently requires substantial contextual interpretation, which is both interaction and collaboration intensive. This effort is bound with the capability, experience and breadth of knowledge of the designer.
- A guideline is a consolidated statement depicting existing design wisdom for which supporting evidence exists. This wisdom ranges from common sense or practical experience to experimentally validated results.
- Guidelines are difficult to communicate to developers since they are not always comprehensible and not correctly targeted to people who should use them in development.
- Guidelines are expressed in very different ways, thus impeding their deployment and correct use by developers. Almost every organization tends to prefer its own proprietary format, thus reducing reusing from one organization to another and breaking any incremental design.
- There is no general agreement about which web design guidelines are correct. [Ratner et al. 1996] conducted a study in which HTML style guides were characterized, compared to established HCI style guides, and evaluated against findings from HCI reviews of web pages and applications. Findings showed little consistency among the 21 HTML style guides assessed, with 75% of recommendations appearing in only one style guide. While there was some overlap, only 20% of HTML relevant recommendations from established style guides were found in HTML style guides. HTML style guides emphasized common look and feel, information display, and navigation issues with little mention of many issues prominent in established style guides such as help, message boxes and data entry. This difference is reinforced by other results showing that HTML style guides addressed concerns of web information content pages with much greater success than web-based applications. It is concluded that while the WWW represents a unique HCI environment, development of HTML style guides has been less rigorous, with issues associated with web-based applications largely ignored.
- Given the steady growth in new sites [Internet Software Consortium 2001], and a severe shortage of UI professionals to ensure usable sites [Nielsen 1999], tools and methodologies are needed to accelerate and improve the Web sites design process.

To address the above issues, researchers have attempted to provide designers with tools for evaluating guidelines. The efficiency of these tools varies according to their purpose (evaluation only, evaluation and automatic repair, evaluation and suggestion of manual correction, etc.) and to the set of evaluated guidelines, but they all have the same functioning principle: detect U&A problems by analyzing the HTML code of the targeted page or other resources like log files to verify some predefined evaluation conditions.

The main shortcoming of these tools is that guidelines' evaluation logic is hard-coded in their evaluation engine. This leads to many limitations:

- Web guidelines are evolving with the continuing technology evolution, so, adding new guidelines or modifying the existing evaluation logic becomes difficult and necessitates modifying and recompiling the evaluation engine even for miner code modification.

- The variety of web user stereotypes, their navigation purposes, their navigation tools, etc. makes it very probable that they do not have the same usability constraints, so, for many of them a site could be considered satisfying a guideline even if it does so partially. Hard coding makes it impossible to specify such evaluation constraints.
- Everything is based on the assumption according to which a web site satisfying a reasonably large set of guidelines is usable. This subsumes that these guidelines are empirically valid.
- It is generally desired that web sites be conform to as many U&A guidelines as possible (generally issued from different sources). For example, it is recommended at the moment that Web sites be conform to W3C WCAG1.0 guidelines [W3C 1999] and to Section508 guidelines (http://www.section508.gov/). Hard coding each of these sets in a separate tool complicates this task. In fact, some existing tools recently integrated both sets in their evaluation engine, but they still treat them separately.

## 1.3 Our Approach to Automating Web U&A Evaluation

Our approach consists in proposing a methodology to improve automated evaluation of U&A of websites by static analysis of HTML code of their pages using the evaluation technique called guideline review.

By *static analysis*, we mean the analysis of the HTML content of the targeted web page as it is seen by the end user by means of traditional browsers, thus, this analysis is supposed to be done after the execution of any potential dynamic content (Javascript, JSP, ASP, etc.), or by ignoring such content if the evaluation is conducted at design phase.

To propose a methodology, we adopt the definition of [Bodart and Pigneur 1989]: A methodology is articulated around four poles:

- The definition of a process made up of steps: the total execution of these steps makes it possible to reach the objective of the methodology;
- The definition of models on which the process is based: modeling hopes to capture the specifications that are essential to the realization of the objective;
- The implementation of software tools: they make it possible to more easily support the methodology in the course of its process and in processing its models;
- The development of an organizational plan (p. e.g. financial means, organizational structure) intended to set up methodology.

The proposal of a methodology implies the study of the first three poles; the latest is related to management [Vanderdonckt 1997].

In concordance with the definition, this methodology will be articulated around:

- A framework that defines a systematic and consistent way for structuring guidelines in order to enable their automatic evaluation;

- A guideline definition language (GDL) able to express guideline information in a sufficiently rich manner to enable an evaluation engine to perform automated evaluation of any GDL-compliant guideline;
- Tools to support the proposed methodology. The tools are **(1)** a structuring tool to enable the specification and manipulation of guidelines structures (structured following the framework and expressed in a GDL-compliant form) and **(2)** an evaluation tool which uses the guidelines to conduct the evaluation of Web sites.

## 1.4 Context of the proposed work

As the definition of the proposed methodology suggests, its ultimate objective is the development of a web automated evaluation tool. The evaluation targets U&A guidelines in first place, but we will see later that we could use such a tool to evaluate guidelines related to other aspects if their evaluation is based on static analysis of Web pages' HTML code.

The main characteristic of the methodology is the separation between guidelines and the evaluation engine to enable their modification at any time. Thus, the user of the structuring tool must be an evaluator with very good HTML knowledge. This also means that the efficiency of the evaluation could vary from a user to another.

It is worth noting that our approach is strictly limited to using the knowledge provided by the content of HTML elements (tags and attributes) in the evaluated pages. Other sources of information that can be obtained from a Web UI for U&A evaluation like events generated during the interaction with the site [Hilbert and Redmiles 2000], log files [Paternò & Paganelli 2001], etc. are totally ignored.

## 1.5 Aim of the research

Appropriate web site design and evaluation methods help to ensure that websites are usable [Nielsen & Mack 1994], but they are so numerous and hard to differentiate which makes identifying which ones are suitable for a particular website a challenging task. This depends on several factors influencing the tool choice: be cost-effective, used for remote evaluation, for real-time evaluation, etc.

In literature, guidelines and evaluation methods for U&A of websites are given and discussed separately. We would like to combine both ideas and even more, i.e. enabling the evaluation of web U&A guidelines by the same evaluation tool.

Our contribution mainly focuses on proposing a methodology to overcome major shortcomings of existing automated web U&A evaluation tools based on the technique of guideline review.

A potential consequence of applying the proposed methodology is the creation of XML-compliant database of formally structured web U&A guidelines. This database would be the kernel of a larger and distributed database of GDL-compliant guidelines if the GDL proves to be suitable enough to enable the specification of guidelines in a formal manner oriented towards automated-evaluation.

## 1.6 Thesis content

As background and motivation for the methodology presented in this dissertation, **Chapter 2** presents a survey of automatic Web evaluation methods and tools.

**Chapter 3** proposes a new methodology for automating U&A evaluation based on separating guidelines evaluation logic from the evaluation engine and dividing the evaluation process into three independent phases: guideline structuring that enable us to specify the evaluation logic, parsing that enable us to capture information needed for evaluation from the targeted web page, and evaluation of the captured U&A data.

**Chapter 4** presents the first pole of this methodology: a framework for structuring web U&A guidelines towards automatic evaluation. The framework aims to provide a systematic and standard manner to structure guidelines in order to identify potential semantic similarities among them.

**Chapter 5** gives a detailed description of the second pole of this methodology: a Guideline Definition Language (GDL) and its syntax and semantics. The presented version of the grammar aims to prove the feasibility of the methodology; it enables the specification of automated-oriented evaluation logic for simple guidelines, but we demonstrate its extensibility to specify more complex guidelines.

In **Chapter 6** we focus our attention on the tools we intend to develop as a support for the proposed methodology. We precise the requirements of them, describe their global architecture, and describe a possible implementation of this architecture. The development of these tools is not finished yet. We conclude the chapter by applying the proposed methodology on a case study. The purpose is to validate its advantages and to identify its limitations.

In **Chapter 7** present a simple cost-benefit analysis: we discuss the different aspects related to the costs and benefits of using the proposed GDL and the KWARESMI evaluation tool.

**Chapter 8** concludes this dissertation. We examine if we reached our objectives, and we underline the advantages and the limitations of the proposed methodology. Potential ameliorations of the methodology are also identified for future research.

All over this dissertation, we apply our methodology on the following relatively complex Web usability guideline: "Select colors that will make your page easy to read by people with color blindness" [Vanderheiden et al. 1997]. In particular, we will use this example to illustrate and validate the content of chapters 4 and 5.