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 3

A Methodology for Automating Web Evaluation

3.1 Introduction

As discussed in Chapter 2, automatic WU&AE tools are promising complements to non-automated methods, such as heuristic evaluation and usability testing. They enable an evaluator to identify potential problems quickly and inexpensively compared to non-automated methods and can decrease the overall cost of the evaluation phase [John and Kieras 1996; Nielsen 1993].

As we mentioned at the end of the previous chapter, our methodology aims to overcome major shortcomings of existing WU&AE tools using the technique of guideline review.

We saw that hard coding the evaluation logic inside the evaluation engine of these tools is the main reason for their inflexibility in facing the urging need to follow continuously changing usability guidelines in the internet world. So, the first sub-objective of our methodology is to separate guidelines and their evaluation logic from the supporting evaluation tool.

We noticed also that it is beneficial to have a mechanism to identify semantic similarities among guidelines issued from different sources. Such a mechanism is absent until now. Our proposed solution for this problem is a framework composed of steps to structure guidelines in a systematical and consistent manner that identifies a maximum of these semantic similarities.

In this chapter we give a global view of our methodology, and the subsequent two chapters will detail the framework and the guideline definition language (GDL). First, this chapter presents two evaluation scenarios as motivation for the methodology.

3.2 Evaluation scenarios

3.2.1 Web site evaluation

As background for the methodology presented in this chapter, Figure 3.1 depicts an evaluation scenario: a Web evaluator seeking to determine the quality of a Web site against a predefined set of Web usability guidelines. If the guidelines have already been structured, the evaluator could use the corresponding XML-Compliant structures as input to a parsing tool. The parsing tool would then scan pages within the site and captures evaluation data corresponding to XML structures. This data is then passed to an evaluation tool that runs the evaluation logic specified in the XML structures and generates an evaluation report. Ideally, the evaluation tool generates a report containing usability errors, their position in

the pages (line, colon) and an explanation as clear as possible of the error. The parsing and evaluation steps could be iterated as necessary using different sets of parsing and evaluation parameters.

As the figure shows, one of the expected advantages of separating evaluation logic from the evaluation tool is the ability to use guidelines from different sources – as soon as these guidelines are put under GDL-compliant formal form-. The sources could be: the local database associated with the evaluation environment (Set2), a custom guideline not yet structured as a GDL-compliant formal guideline, and even remote GDL-compliant guidelines (Set1).

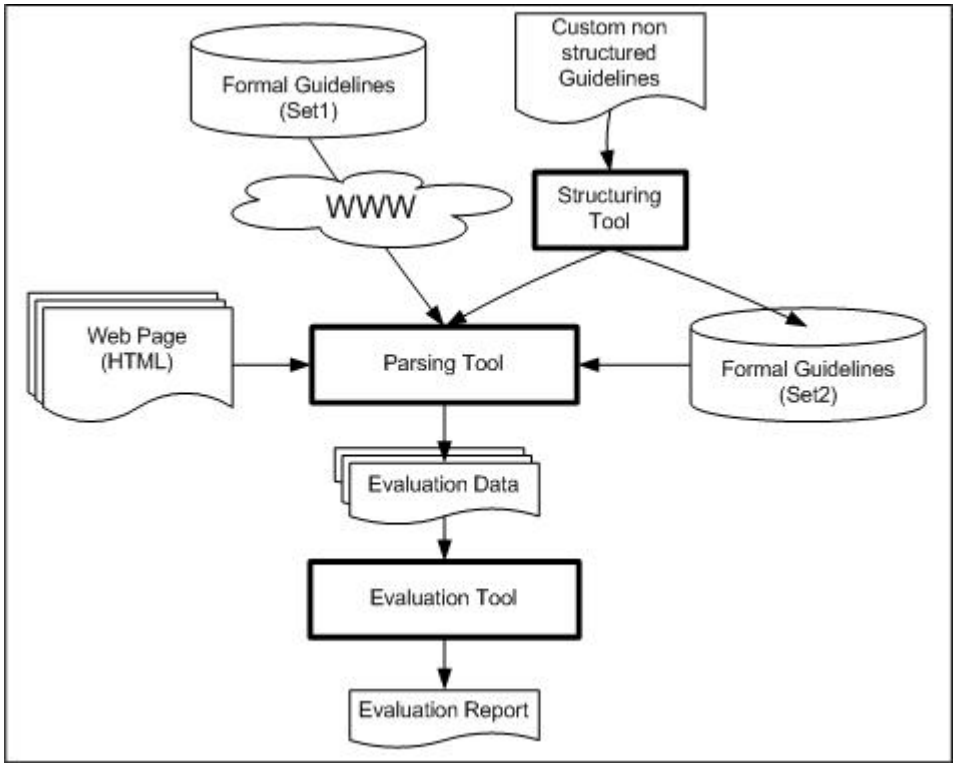

Figure 3.1: A Web site evaluation scenario. Using XML to specify structures makes it possible to import guidelines from remote sources if these guidelines are stored as XML files on the remote machine.

3.2.2 Web page design

One can also imagine that a Web designer would want to obtain guidance on web designs earlier during the design phase as opposed to after implementation. As we will see later in chapter 4, a guideline structure provides some usability-related information and detailed HTML-related information about the guideline. If a tool for managing XML structures is implemented, then it would be possible to use it to select, from the structured ones, those targeting a given usability concept (like navigation) or ability property (like vision) or dealing with a given kind of Web concepts (like images, colors, etc). Scapin et al. [2000] presents a useful and relevant framework for organizing Web guidelines that includes a taxonomy of index keys (e.g., alignment, buttons, downloading, headings, language, scrolling, navigation structure, and so on); this taxonomy could be used to index the XML structures.

3.3 The Methodology

Figure 3.2 depicts the methodology developed to support the Web site usability and accessibility evaluation scenario.

The methodology is based on decomposing the whole evaluation process into two distinct but related phases: guidelines structuring and Web evaluation. The evaluation phase consists of two sub-phases: parsing targeted Web page and conducting evaluation. The two main phases are totally independent, which gives many improvement possibilities at each of them.

We must precise that we target pure HTML only (no CSS, scripting languages, etc.). We will demonstrate in chapter 5 that the proposed methodology can be extended to pages using CSS.

3.3.1 Guideline Structuring

The first phase consists of structuring guidelines in terms of HTML elements. Obviously, this activity is highly influenced by the understanding of the original guideline semantics. In addition, it requires good HTML knowledge to identify tags and attributes that can be used to conduct the evaluation.

Guideline structuring aims to transform the highly abstract human knowledge easily expressed in the guidelines using natural language into a concrete knowledge understandable by the UE tool. The interface between these two levels is the GDL (see figure 3.3).

The structures are stored under XML form that respects the syntax of the GDL. Using XML format enables us to profit from the many advantages of XML: structures can be manipulated easily by any text editor; we have many tools to check structures for well-formedness and for validity compared to the predefined GDL DTD, we can store structures in a database or as XML files, etc.

Does separating guidelines from the evaluation tool gives this tool the desired flexibility in following the rapid usability and accessibility guidelines' evolution? We think that the answer is YES as soon as the underlying GDL is rich enough to enable the expression of the evolution logic needed to evaluate the new guidelines. In fact, from its name, GDL is a specification language, so, its syntax can be extended to follow this evolution. As guidelines are independent from the evaluation tool, their logic can then be modified according to the new GDL syntax and this modification will be transparent for the evaluation tool.

It is worth noting that, according to the proposed methodology, the structuring phase is totally independent from the evaluation phase. What the later phase needs to know to parse the targeted page and conduct its evaluation will be included in the guidelines structures or provided by the evaluator as parameters that are mainly intended to control the parsing and evaluation processes.

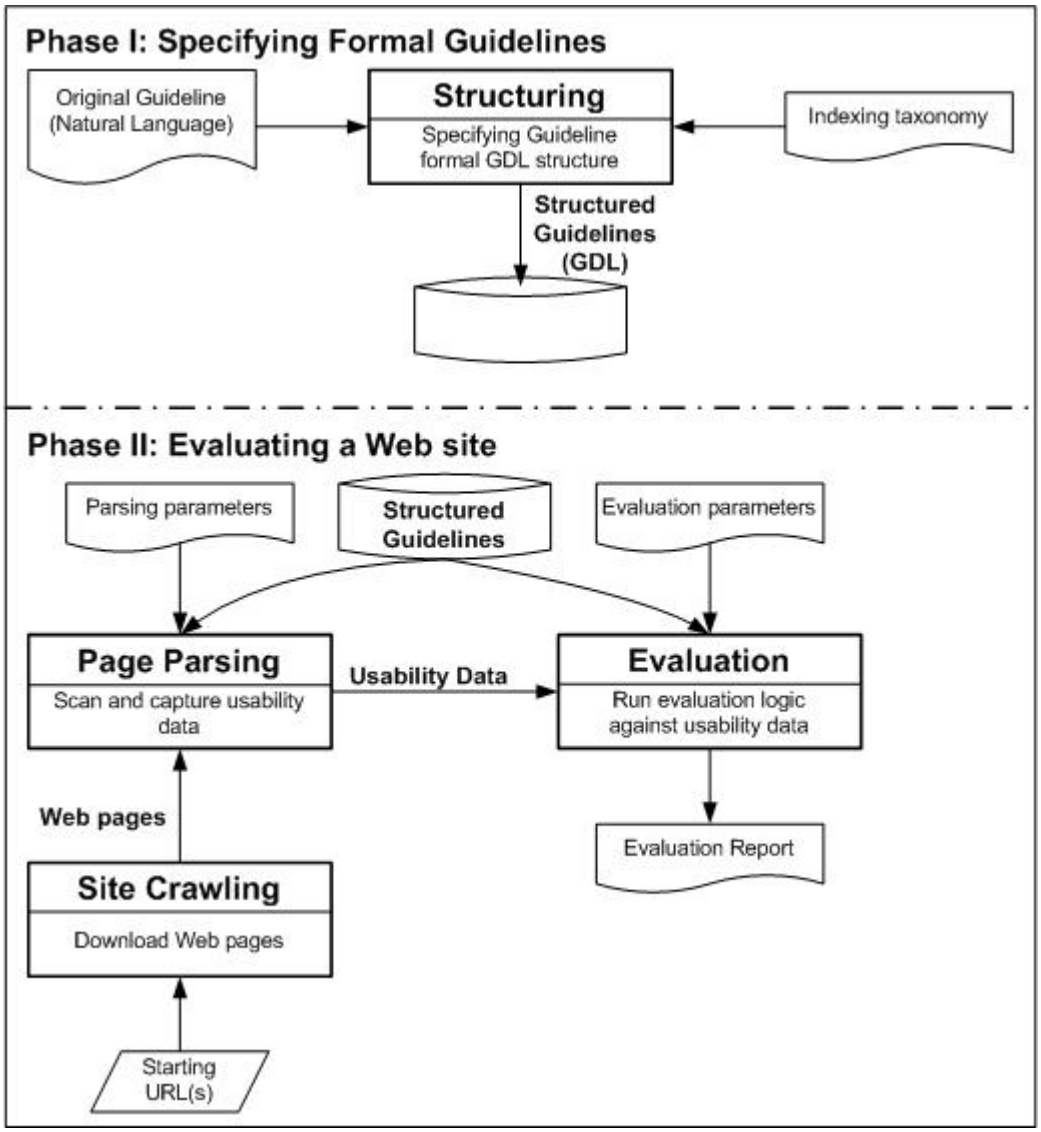

Figure 3.2: Overview of the steps of the proposed methodology. The two phases are totally independent. Parsing a Web page is leaded by some parsing parameters that the evaluator can manipulate (like conditions on stopping the parsing process), but also by some information included by the guidelines structure (like what usability data to capture in the parsed Web page).

3.3.2 Web page parsing

The parsing process is based on the information provided by the structures of the targeted guidelines. As we will see in chapter 4, the structure of a guideline provides detailed information about:

- Usability indexing information to help organizing guidelines according to an indexing taxonomy.
- The HTML elements that must be examined (if exist) in the evaluated Web page.
- The priority level of each HTML element.
- The relationships among considered HTML elements, some of these relationships could be useful in improving the parsing process.
- The evaluation logic that must be applied on these HTML elements to check the reviewed guidelines.

All these information can be used to control and improve the parsing process. In addition, we can specify many parsing parameters to control this process: which guidelines to evaluate, number of desired instances (<=N, all), possibility to ignore some HTML elements during the parsing, etc.

As figure 3.2 shows, if we want to evaluate an online Web site, we start by downloading its pages and storing them locally. Note that the parsing and evaluation are conducted on the downloaded pages individually because the proposed methodology supports evaluation at page level only.

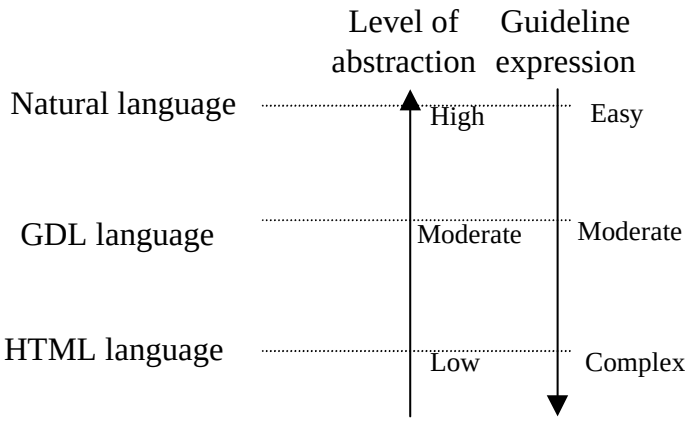

Figure 3.3: Levels of abstraction in the proposed methodology

3.3.3 Guideline evaluation

After parsing the web page, we can apply the evaluation conditions that we defined during the structuring phase on the captured usability data. Every condition is applied on the captured instances of its corresponding HTML elements to determine respect or violation of the guidelines. By this way, a detailed evaluation report can be generated on respected/violated elements, number of detected instances, percent of respect/violation, etc.

3.4 Evaluation Improvement

Decomposing the evaluation process as described above offers improvement possibilities at each of the process phases.

3.4.1 Guideline Structuring

As parsing Web pages is based on evaluation sets[1] and evaluation conditions defined in this phase, we can improve the evaluation at two levels: for a single guideline, there are two ways: identifying the minimum ensemble of sets needed to evaluate the targeted guideline, and expressing conditions in the most forward way to minimize the number of operations that evaluation engine would need to execute them.

At the level of many guidelines, we can improve evaluation by identifying common structures or sub-structures. This improvement cannot be neglected since guidelines are expressed at a high abstraction level, and as they come from

[1] An evaluation set is an ensemble of HTML elements (see next chapter)

different sources, it is very possible to have guidelines that are totally or partially semantically identical.

3.4.2 Web page Parsing

The first significant improvement during this phase is the use of the concept of exclusion among evaluation sets. By definition, one evaluation set Excludes one (many) other evaluation set(s), in a given evaluation context, if its presence excludes its (their) evaluation. This concept is based on the *Scope* concept related to HTML elements. Generally, the excluding set has an element whose scope is within the scope of an element of the excluded set. Of course, these two elements must have the same rendering effect. For example (Fig. 3.4), in the context of text color evaluation, a set containing the attribute Table.bgcolor (like S1={Body.text, Table.bgcolor}) excludes a set containing the attribute Body.bgcolor (like S2={Body.bgcolor, Body.text}), because the scope of Table.bgcolor overcomes the scope of Body.bgcolor.

The second improvement is to combine parsing and evaluation in one step. This means that an evaluation condition is triggered as soon as an instance of the associated evaluation set is completely detected in the evaluated Web page. In this case, we can specify some improvement parameters such as stopping the evaluation if we detect a given number of violations.

This combination of steps would be optional because, in some situations like the need for a detailed evaluation report or to repair ergonomic problems, it is desired to capture all instances of evaluation sets (even non completed or negative ones).

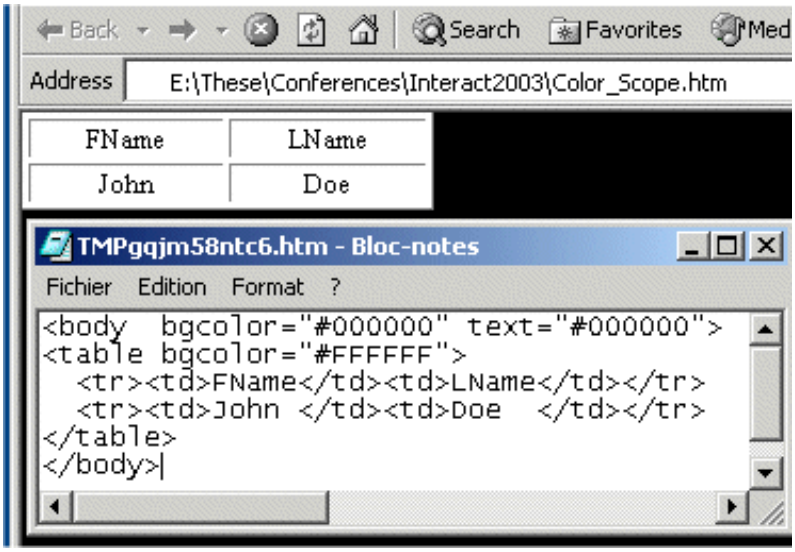

Figure 3.4: Scope of Table.bgcolor is within the scope of Body.bgcolor

3.4.3 Guideline evaluation

The improvement that can be done during this phase relies mainly on improving the execution of evaluation conditions. We will see in the next chapter that we introduced many concepts to facilitate the identification of similar or identical parts among evaluation conditions. The use of these concepts would allow the evaluation tool to minimize the number of executed operations to conduct the evaluation of predefined guidelines against captured usability data. This

improvement would have significant impact when evaluating large or very large Web sites.

3.5 Evaluation activity based on the proposed methodology

In this section we will analyze and discuss the activity of automatically evaluating a web site according to the proposed methodology. Then, we will try to identify its limits.

3.5.1 Tasks

Figure 3.5 depicts that tasks and sub-task that must be accomplished by a tool based on our methodology to conduct assisted or automated evaluation:

- **Guidelines preparation**: this task consists in *coarse selecting guidelines* from their sources (W3C, Section508, ISO, etc.). If it is the first time of using a guideline, it must at first place be *structured under a GDL compliant form*. This selection is conducted manually for new guidelines because guideline sources are not always accessible in the same way (Web sites, books, articles, etc.), and they are generally expressed in a natural language informal form. For most existing evaluation tools, this task is done only once when starting the development of the tool because guidelines will be incorporated inside it, but in our case, it can be repeated more than once because the tool is independent of guidelines.
- **Evaluation preparation**: this task consists in *selecting individual guidelines* related to the ergonomic aspects targeted by the evaluation (accessibility, usability, user satisfaction, etc.). Next is the *configuration of capturing and analysis phases* of the evaluation with the help of related configuration parameters. The assisted evaluation configuration is generally conducted by the evaluator at every evaluation session. It can be automated in some circumstances. For example, the evaluation configuration can be saved in a database or a configuration file that will be automatically read by the evaluation tool.

In fact, it is possible to combine our automated guideline review evaluation with other evaluation techniques like, for example, configuring the capturing phase in a cognitive walkthrough-like way. In some cases, like repeated evaluation of the same Web site, we can specify an evaluation path: {evaluate page P1, then Pi}, {Evaluate Pi and linked pages, etc.} or specify some conditions on evaluated guidelines: {evaluate guidelines Gi, Gj IF Gk is Ok, evaluate Gi on Objects Om and On, etc.}.

- **Evaluation execution**: when everything is ready, the tool automatically parses the Web pages to *capture* needed data, then it *analyses* them to check if they respect/violate the targeted guidelines. At the end of the analysis, the tool *generates* and evaluation results. Depending on the site's size and/or other evaluation constraints, the evaluation covers the whole site or a selected set of its pages.

- **Evaluation Results visualization**: the final task of the evaluator is to examine the generated results in order to take needed actions. These results can be saved for future use or deleted by the end of the evaluation session.

3.5.2 Quality factors

The quality of the evaluation produced by the proposed methodology depends among others on the following factors:

- **Intrinsic quality of the guidelines**: as our objective is to enable a dynamic and evolving manipulation of Web guidelines issued from different sources, the quality of the introduced guidelines (especially those issued from non established sources) influences the global quality of the tool. It is very important to choose guidelines that have real impact of the ergonomic quality of a Web site.
- **Guidelines application**: a critical quality factor of the methodology is the well interpretation and formulation of a guideline GDL compliant form. The evaluator's awareness of the guideline aims, his HTML knowledge and his GDL experience are the main elements that determine the quality of this application.
- **Implemented GDL features**: one of the biggest challenges that we must face is the ability to provide a rich set of GDL constructs to enable an easy, complete and flexible Structuring, and in the same time implement all these constructs in the tool.

3.5.3 Automation Limits

We can identify two kinds of limits:

- **Limits related to the technique of Guideline Review**: they we will be automatically reflected in any evaluation tool based on it. As mentioned in 3.1.1, some tasks –like guidelines collection- must be done manually.
- **Limits related to our methodology**: they will be reflected in our tool but we can still hope they will disappear with the evolution of the methodology. The main limit of the methodology resides in the incorporation of guidelines logic inside the evaluation tool. In our case, this incorporation is not direct because the evaluation logic is expressed via the GDL outside the tool. The use of guidelines is done by interpreting the GDL expressions, which means that it depends on the quality of the interpreter and on the richness of the GDL.

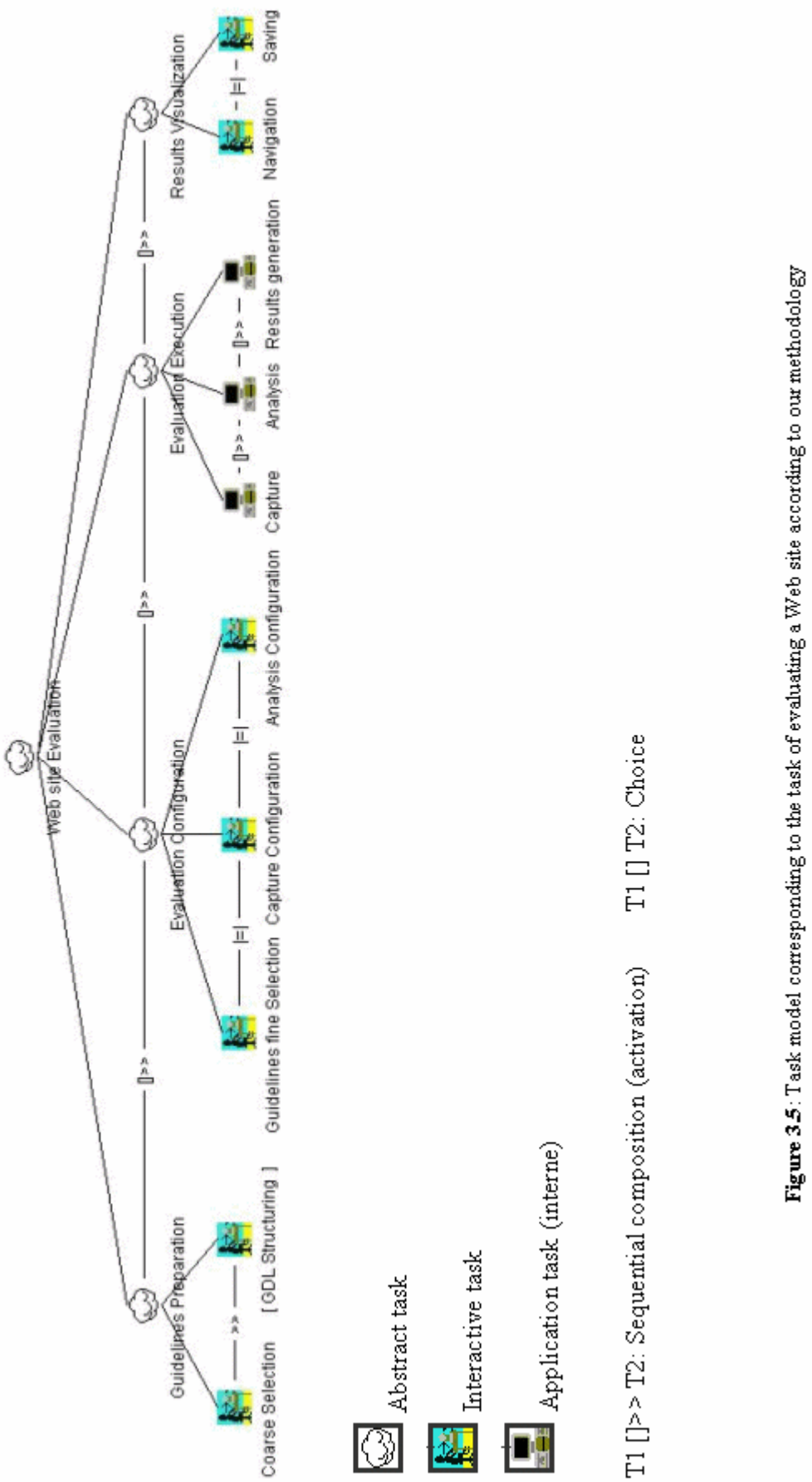

Figure 3.5: Task model corresponding to the task of evaluating a Web site according to our methodology

3.6 Development of automated Web evaluation tool

Our ultimate objective is to develop an automated Web evaluation tool that can be used to evaluate Web-related guidelines (usability, accessibility, etc.).

3.6.1 Developing a Tool Based on Guidelines

To develop any tool for working with guidelines, five development milestones have been identified for reference and comparison purposes [Vanderdonckt 1999], but also for structuring the process to reach that goal. The five development milestones, through which one must pass to produce a high quality tool for working with guidelines, are the following:

A) Guidelines collection

The goal of this first step is to gather a useful subset of guidelines suitable for designing Web pages. An initial unstructured but comprehensive set of guidelines is formed by collecting, gathering, merging, compiling guidelines from all available world- wide ergonomic sources. These different sources are not especially dedicated to Web usability (i.e., they focus more on graphical user interface in general), thus requiring some modification, adaptation, extension, and so on.

B) Guidelines organization

In this step, because the initial set of guidelines is copious and ranges over many evels of rigor and credibility, guidelines are classified in a good organizational structure. Since the initial set, guidelines are organized proceeding by two activities:

- **Classifying each guideline by ergonomic criteria**: An ergonomic criterion is hereby defined as a well- recognized usability dimension in human- computer interaction whose reliability effectiveness and impact on usability have been experimentally assessed. Each guideline has therefore been classified by a sole ergonomic criterion based on its definition. By this classification designers have a first idea on when and where the related guideline can be applied as well as some first idea of its absolute level of importance.
- **Further classifying each guideline by alternate index keys**: as the set resulting by first step is still wide, guidelines are further classified by alternate index keys. This classification allows multiple and flexible access paths to each guideline, rather than merely by ergonomic criteria [Bastien et al. 1999]. Such accesses permit automatic identification of a certain guideline, so that it can be evaluated more rapidly (e.g., Navigation structure, Links/Organization, etc). Each guideline is then assigned to one or many methods and techniques for Web site design/evaluation that can be effectively used to assess the guideline (i.e., to indicate which level can be used for automation). Moreover, to each guideline can be attached a score in order to express the impact of a violated or respected guideline.

C) Incorporation of guidelines into approach

It is now important to associate sections of guidelines with the different phases of a development life cycle.

In this way, it is expected that sets of guidelines suitable for each phase can be more easily identified and accessed. Thus, the goal of this step is to locate points within these phases where organized guidelines should be considered, to specify which should be considered by identifying local guidelines (for a phase), global guidelines (for all phases) or pervasive ones (for several continuous phases).

D) Operationalization

Guidelines incorporated into a development life cycle are typically used manually. To embody them in a software tool, a further stage of guidelines operationalization is thus aimed at re-expressing them in a more formal way, so that it results easier to evaluate them automatically.

E) Guidelines use

In our case, the guidelines will be used to conduct a Web evaluation.

3.6.2 Our Tools

The main objective of our work is to propose a methodology for automated evaluation of Websites by guideline review. In order to accomplish this task, we need to accomplish the tasks corresponding to the steps motioned above. Accomplishing all these tasks by one tool is not trivial and could affect the efficacy of the evaluation process, especially if we target a flexible and efficient tool. In addition, these tasks can be done separately. For these reasons, an ideal solution is to develop separate tools to accomplish these tasks, and integrate them within a system in order to conduct the automated evaluation.

Thus, the tools that we should be developed to support the methodology are:

1) A management tool of ergonomic knowledge

This module manages the ergonomic knowledge bases at various levels: insertion of the new guidelines in the system database, divided and collaborative edition of the existing guidelines (it should be possible to enrich the database by anyone via the Web), selection of the guidelines corresponding to a given context (targeted user stereotypes, type of site, types of tasks, etc.). In addition to management information about a guideline (source, indexing keys, comments, etc.), a field will contain the specification of this guideline in GDL-compliant form.

2) A tool for structuring guidelines (GDL editor)

This tool will make it possible to formally specify a guideline in GDL-compliant form and to store this specification in the database or in an XML file to be exploited later by the evaluation tool.

3) An automatic Web Evaluation Tool

On the basis of some evaluation parameters, this module evaluates the ergonomic quality of a page, a series of pages or a whole site by subjecting it to a set of ergonomic guidelines taken from the databases or XML files. It produces a

customizable evaluation report. The pages having ergonomic problems are isolated to be treated by the ergonomic reparation tool. The evaluation tool cannot obviously automate the evaluation of all the guidelines in a complete way (the formal GDL specification will provide necessary information indicating their level of automation: partial, total, with a percentage).

Remark

We can see that the tools described above constitute a global system for working with guidelines respecting the milestones of 3.5.1 where:

- The management tool enables guidelines collection and organization (A, B);
- The structuring tool enables guidelines operationalization (D);
- The evaluation tool enables guidelines use (E);
- The system composed by the integration of these tools enables the incorporation of guidelines into a Web site life cycle (conception, evaluation, reparation) (C).

3.7 Summary

This chapter presented global overview of a Web evaluation methodology consistent with evaluation methods based on the technique of guideline review. Unlike other Web usability and accessibility evaluation approaches, this approach uses a formal language to specify guidelines evaluation logic outside the evaluation tool.

What distinguishes this approach from other guideline review methods is:

- The separation between guidelines and the evaluation tool to enable dynamic and flexible manipulation of evaluated guidelines; and
- The use of a formal language to re-express guidelines in a formal and structured manner in order to improve the automation of their evaluation. In addition, the language is XML-compliant, so, this should enable a kind of portability of the structured guidelines and the evaluation tool.

The methodology's ultimate objective is to define a good basis to implement a WU&AE tool. The pillars of this methodology will be discussed in details in the following chapters.