The Impact of Accessibility Assessment in Macro Scale Universal Usability Studies of the Web

Rui Lopes
LaSIGE/University of Lisbon
Campo Grande, Edíficio C6
1749-016 Lisboa, Portugal
rlopes@di.fc.ul.pt

Luís Carriço
LaSIGE/University of Lisbon
Campo Grande, Edíficio C6
1749-016 Lisboa, Portugal
lmc@di.fc.ul.pt

ABSTRACT

This paper presents a modelling framework, *Web Interaction Environments*, to express the synergies and differences of audiences, in order to study universal usability of the Web. Based on this framework, we have expressed the implicit model of WCAG and developed an experimental study to assess the Web accessibility quality of Wikipedia at a macro scale. This has resulted on finding out that template mechanisms such as those provided by Wikipedia lower the burden of producing accessible contents, but provide no guarantee that hyperlinking to external websites maintain accessibility quality. We discuss the black-boxed nature of guidelines such as WCAG and how formalising audiences helps leveraging universal usability studies of the Web at macro scales.

Categories and Subject Descriptors

H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*User issues*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/ methodology*

General Terms

Design, Human Factors, Measurement.

Keywords

Audience Modelling, Universal Usability, Web Interaction Environments, WIE

1. INTRODUCTION

The growth of the Web is profoundly changing the way people interact with information and with other people. This has led to an expansion of opportunities for the Web on different vectors, including the massive production of contents for all by all (e.g., Wikipedia[1]). To cope with this growth, Schneiderman proposed a research agenda for Universal Usability [27], one of the key challenges of Web Science [2, 29]. This challenge is centred on exploring the accommodation of software (and their user interfaces, in particular) to both technology variety, user diversity (including universal access issues) and knowledge, instead of providing an one-size-fits-all solution to everyone.

The increase of population accessing the Web has brought attention to specific sectors that should not be ignored, such as the elderly, children, the disabled, etc. The availability of Web browsers on devices other than the PC (e.g., mobile phones) has also contributed to this expansion of the Web. Combining these factors and particular usage situations (e.g., public spaces) raises new challenges. Moreover, new expectations have brought different user intentions (e.g., finding information *vs.* entertainment), leading to richer Web interaction requirements that must be met, as different goals might require different interaction metaphors.

However, when designing and implementing a website, especially the interactive side of it, it is laborious to maintain high standards of usability across the whole spectrum of audiences. Consequently, some audiences are usually ignored, as it is unfeasible to create a different version of a Web front-end to each of them, often resulting on severe accessibility and usability issues. Moreover, when scaling to multiple webpages, users are typically faced with different levels of usability and accessibility (either intra or inter website), which further breaks their user experience.

To mitigate such issues, experts have defined several universal usability guidelines that meet the expectations of users, such as WCAG [4], ISO/TS 16071:2003 [12], etc. Traditionally, such procedures require either expert analysis to be performed (with or without real users in the process), in order to yield a quality mark for the audience they are targeted to. This poses severe difficulties on performing large scale studies of universal usability of the Web. Consequently, there still is much latent information about universal usability that is hidden on the Web (e.g., at what quality level are users supported by websites, and how it influences their user experience).

In this paper we present a modelling methodology to express Web Interaction Environments (WIEs), facilitating the study of universal usability issues in different fronts. We use this methodology to leverage the implicit audience inherent of WCAG 1.0, and conduct a large scale study of Wikipedia's accessibility quality. We end the paper with a discussion on the limitations of automated universal usability studies based on the limitations of such guidelines.

[1] http://www.wikipedia.org

W4A2008 - Technical, April 21-22, 2008, Beijing, China. Co-Located with the 17th International World Wide Web Conference.

2. RELATED WORK

When assessing the universality of usability of the Web, several factors should be taken into account from the beginning. First, there is a need to understand what audiences are being analysed, how to characterise them, and what similarities and differences can be found between them. Second, due to the large scale factor of assessing universal usability quality, manual inspection procedures cannot be used. Therefore, quantification procedures must be automated. Lastly, different patterns might be analysed, based on both audience modelling and universal usability quantification processes. Next, the current practices for these three topics are further analysed.

2.1 Audience Modelling

One way to look at audience modelling is the characterisation of users and devices by modelling corresponding multimodal interaction. In [21], the authors present a UML framework to express multimodal interaction in terms of the sensory, perceptual, and cognitive effects they produce. It has been applied in different contexts, such as modelling universal accessibility [20] and content repurposing [22]. This framework is mostly targeted to researchers that need to describe multimodal scenarios in-depth, such as describing what human entities are required for speech-interaction tasks. However, by being low-level, it becomes too cumbersome to simply describe audiences and to explore their synergies and differences.

Audience modelling is akin to context-awareness and context-sensitivity. The notion of context encompasses personalisation features (i.e., tailoring to users), device tailoring (e.g., mobile devices), and information-sensing (e.g., GPS). Several Web engineering practices have been using context to design adaptive Web applications, including [32, 8, 5]. This type of practices afford the exploration of already existing Web applications [9], as well as inferring context-awareness capabilities. However, this is a non-trivial task, since they are mostly tied to enhancing application modelling tasks at highly abstracted levels (data models, navigation, presentation structures) [15], not evaluating universal usability of Web front-ends.

Ubiquitous user modelling [11] also provides some clues for tailoring Web front-ends, especially on content personalisation scenarios. Here, models express individuals and devices characteristics, actions, and navigation history to adequate content more effectively. Despite having a different goal, ubiquitous user modelling provides insightful cues to which characteristics should be studied to express WIEs effectively. In [10], a thorough discussion presents different methods to obtain these models automatically.

The growth of interest for universal usability has provided interesting knowledge for the study of Web Interaction environments. In [28], a multi-layer interface design is proposed to enable a gradual use of complex systems, depending on user expertise, in order to reduce knowledge gaps [1]. We believe that this type of approaches can be further extended and woven into audience modelling practices.

2.2 Quantification of Universal Usability

Assessing the level of usability of a user interface is a critical measure for the success of software applications [25], including websites and Web applications [19]. As these methodologies become cumbersome to verify (typically require manual inspection by experts), it is often left aside by Web designers [14, 23]. Furthermore, as content production and dissemination on the Web becomes deeply democratic, automated approaches become of vital importance.

To mitigate such problems, automatic procedures should effectively assess and quantify how accessible and how usable websites are, especially to each audience. In [13] the authors present a state-of-the-art survey of a set of procedures that afford the automatic quantification of usability. However, these methodologies and frameworks often assume audiences as being heterogeneous, yet evaluated from a homogenous point of view. To mitigate such problems, evaluations have to be tailored to user characteristics, as well as the device they are interacting with [35], instantiated on a set of quantitative metrics for assessing Web accessibility quality [34].

2.3 Large Scale Web Studies

Universal usability studies can only be scaled to the dimension of the Web based on the understanding of user requirements and being able to assess if a website supports them (i.e., through quantitative metrics and automated approaches). While large scale studies of the Web have been conducted several times, as surveyed in [7], little or none is known about the impact of the universal usability of a single website, in the context of the Web as a whole. While some answers on structural analysis of websites have been given before [3], they do not convey information about their adequacy to the real world of high user diversity.

3. WEB INTERACTION ENVIRONMENTS

Studying the universality of the usability of a Web front-end should start first by understanding which audiences are required to be supported and, afterwards, explore their similarities and differences. The concept of Web Interaction Environments (WIE) emerges from this scenario. A WIE is defined as a particular audience's group of intrinsic characteristics upon which tailored evaluation procedures are applied to a website. Websites can support more than one WIE, and characteristics may be shared between them. This way, the synergies between each WIE may be explored.

WIEs should not represent particular instances of groups of characteristics, such as *User John Doe* or *Device FooBar*. Instead, WIEs should focus on aggregating characteristics that represent a specific group (e.g., *Blind* represents all users that have a visual impairment).

Based on the challenges proposed by Schneiderman [27] (i.e., *technology variety*, *user diversity*, and *gaps in user knowledge*), we have devised four characteristics domains:

- *Users*, encompasses intrinsic *user diversity* characteristics (either mutable or immutable);
- *Devices*, provides support for *technology variety*;
- *Usage situations*, affords putting the *User* interacting with a *Device* on miscellaneous situations (accommodating external factors that influence user abilities);
- *User intentions*, bridges the *gaps in user knowledge*, regarding Web front-end interactivity.

In the following sections, each one of the four characteristics domains are described more thoroughly.

3.1 Users

There is a real diversity of users interacting with the Web on a daily basis. This fact becomes visible and more relevant typically due to traffic increase and broadening of the spectrum of interest for a given website. When websites have to cope with different groups of users, while maintaining user experience quality, each group's characteristics must be carefully taken into account. Hence, a WIE should be able to classify and characterise users accordingly.

These requirements can be characterised from different perspectives, such as physical and cognitive (dis)abilities, age, gender, cultural aspects, etc. By selecting appropriate characteristics, WIEs can describe user groups with a large coverage, such as *Elderly* or *Unimpaired*, or highly specific, such as *Colour Blind* or *Partially Sighted*.

3.2 Devices

With the increasing number of internet connected devices, websites can be accessed by users from devices other than a typical PC. This includes PDAs, mobile phones, kiosks, TVs, gaming consoles, etc. Accessibility dependent users have also brought attention to their specialised devices, e.g., braille terminals. Such diversity encompasses a whole range of input and output modalities, as well as different configurations on each one. Hence, it is very unlikely to provide the same usability quality on each device without tailoring Web front-ends appropriately.

In order to understand which devices should be supported by a website, a WIE should express the characteristics of each device group, including its input and output modalities, features, limitations, etc. Therefore, WIEs must be able to describe the device ecosystem supported by a website's front-end, and make explicit the possible synergies between each device group. Examples of device oriented WIEs could be *Keyboard based devices*, *Visual output devices*, or *PDAs*.

3.3 Usage Situations

Another factor to take into account when defining the WIEs of a website, concerns its usage situation. Users cannot be perceived independently from the device they use to interact with the Web. By affording the exploration of synergies between different WIEs, usage situations can bridge *User* and *Device* clusters of characteristics, allowing teams to study the effects of putting the user *using* the device, which deeply influences user experience on Web interaction.

Two distinct usage situations can be studied with WIEs: environment and user/device intertwining. The former relates to situated scenarios, i.e., studying the user with the device on a real world situation. Some characteristics include public spaces, background noise, lighting, intermittent internet connectivity, etc. This allows teams to explore WIEs such as *City* or *Night*, and how it influences the user experience quality that has been delineated.

The latter usage situation, user/device intertwining, is centred on understanding the possible accessibility and usability problems that may appear when users interact with a Web front-end in specific devices. For instance, consider the case where a website must support a WIE for blind people and another for mobile devices. When both WIEs are taken into account at the same time, the inherent user experience quality is degraded, as design or implementation decisions supported by each WIE are incompatible (e.g., inability to see *vs.* a screen display). Another example relates to merging WIEs for unimpaired users and mobile devices. If the user wants to interact with a website on-the-go (e.g., while walking), the UI should be as less intrusive as possible, due to the decrease on the user's attention level.

3.4 User Intentions

The last characteristics domain that can be used in WIEs concerns user intentions. Some websites are targeted to very specific audiences or to serve a single purpose (such as provide information). However, other websites can be used for broader tasks. For instance, considering Amazon.com[2], it might be used for information tasks (information about a book), or to perform transactions.

As each user may have a different intention while interacting with a website through its front-end, the way information is presented and navigable deeply influences its corresponding usability. Consequently, WIEs afford the explicit description of user intentions through a set of core characteristics, that help multidisciplinary Web development teams to adequate user experience to these situations.

4. VOCABULARY

In order to describe the different characteristics that encompass the definition of a WIE, we have created a supporting vocabulary incorporating more than 130 concepts, using OWL [26] and inherent ontology engineering best practices [31]. By having a common and formal vocabulary to characterise WIEs, ambiguity is reduced when discussing interaction scenarios. Moreover, this formalisation opens the way for creating tools that exploit WIE characterisation activities, such as usability assessment engines, expert systems, or even model driven development practices.

In consonance with the described WIE domains, we opted to create an ontology that unifies the corresponding vocabularies of the four domains, as seen on Figure 1. The root concept, *Characteristic*, represents the highest abstraction of all characteristics. Each parent/child relationship on the ontology is defined as *is-a* (e.g., *Autonomy* is a *Device Characteristic*). Consequently, higher depths on the ontology tree represent more specific characteristics.

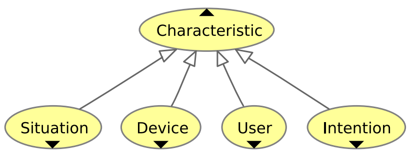

Figure 1: Main vocabularies for WIE characteristics

Regarding the description of user characteristics, several approaches have been taken before. The most important has been taken by the World Health Organization, through ICF[3] (International Classification of Functioning, Disability and Health). ICF describes concepts such as body functions and structures, for impairment qualification on medical diagnosis tasks. We opted not to embed this classification directly on our ontology due to the fact that it lacks concepts that could be helpful on defining and discussing user centred WIEs. For instance, this classification does not enumerate the different types of colour-blindness. Instead, it

[2]http://www.amazon.com

[3]http://www.who.int/classifications/icf/site/index.cfm

describes *Colour vision* as *Seeing functions of differentiating and matching colours*. Nevertheless, this classification provided helpful clues on what user characteristics should be expressed in our ontology. Figure 2 presents some of the concepts extracted from the *Users* domain.

```
User
  UserAbility
    AbilityCognitive
    AbilityPhysical
  UserCharacteristics
    CharacteristicsAge
    CharacteristicsHandedness
  UserCulture
    CultureLanguage
      LanguageWritingDirection
  UserDisability
    DisabilityCognitive
      CognitiveMemory
      CognitiveReading
    DisabilityPhysical
      PhysicalBlind
      PhysicalMotor
```

Figure 2: Partial *User* domain characteristics

Concerning device characteristics, we have defined the main features that are present on different devices, based on categorisations defined elsewhere [36, 18, 11]. However, none of existing categorisations provide an in-depth enumeration of input and output modalities typically found on universal interactivity scenarios on the Web.

Hence, we have defined a *Device* domain that encompasses these specific issues. Modalities might be present on the device as hardware specifications (e.g., visual display), operating system features (e.g., accessibility helpers), or Web browser properties (e.g., accepted media types). An insight on the *Device* domain is presented on Figure 3.

```
Device
  DeviceAutonomy
  DeviceInput
    InputKeyboard
    InputPointing
      PointingMouse
      PointingTouch
  DeviceNetwork
  DeviceOutput
    OutputAural
    OutputHaptic
    OutputVisual
  DeviceUserAgent
    UserAgentMediaAccept
    UserAgentType
```

Figure 3: Partial *Device* domain characteristics

For characterising usage situations in WIEs, its specific domain was enriched with concepts that afford both user situations (e.g., device handling) and ubiquity situations (e.g., connectivity, noise). This has been partially derived from the UbisWorld ontology [11], where different ubiquity concepts are categorised. An excerpt of the ontology that supports the *Usage Situations* domain is presented on Figure 4.

Lastly, we mapped the different user intentions characteristics into a specific domain within the ontology. We have adopted a specific characterisation proposed in [16], where user intentions have been classified through a field

```
Situation
  SituationConnectivity
  SituationEnvironment
    EnvironmentLightning
    EnvironmentNoise
    EnvironmentSpace
      SpacePrivate
      SpacePublic
  SituationHandling
    HandlingDualHanded
  SituationMovement
  SituationPosture
```

Figure 4: Partial *Usage Situations* domain characteristics

study centred on user activity monitoring methodologies. On Figure 5 we present the characteristics for the definition of WIEs based on the *User Intentions* domain.

```
Intention
  IntentionExchange
    ExchangeCommunications
    ExchangeTransactions
  IntentionMaintenance
  IntentionSeeking
    SeekingBrowsing
    SeekingFactFinding
    SeekingInformationGathering
```

Figure 5: *User Intentions* domain characteristics

Next, we present a modelling framework that affords the description of WIEs, encompassing the vocabulary described on this Section.

5. MODELLING FRAMEWORK

The ontology we presented plays a critical role on describing WIEs, by providing a comprehensive set of characteristics. However, listing the concepts is not enough to fully describe WIEs. It should be possible to synthesise them in a simple way, in order to ease the task of exploring the differences and synergies between WIEs.

We have defined a modelling framework supported by the vocabularies presented on the previous Section, supported by a graphical description of WIEs, as depicted on Figure 6. This framework is based on a domain specific meta-model first presented in [17], which supported modelling document interaction scenarios.

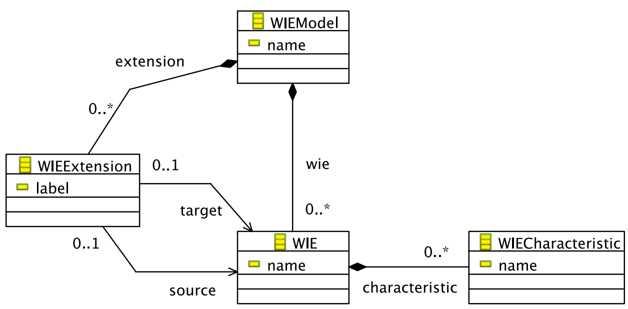

Figure 6: WIE modelling meta-model

The WIE meta-model defines the basic building blocks to help expressing WIEs. At its root, the WIEs model (*WIEModel*) is composed by a set of classes (*WIE*) and their associated characteristics (*WIECharacteristic*). Each *WIE* instance groups a set of *WIECharacteristic* instances, thus explicitly describing a particular interaction environment. Each instance of a *WIECharacteristic* corresponds to a particular concept from the vocabulary presented before.

It is worth mentioning that the modelling framework works on an open world assumption, i.e., what is not stated as a characteristic on a WIE is not necessarily false, it is simply unknown. This feature enforces the explicit nature required to express which factors are taken into account. For instance, not providing a *wie:UserDisability* derived characteristic does not mean that non-impaired users may be associated with that particular WIE. It simply means that the WIE does not explicitly supports the disabled. Consequently, to champion non-impaired users, a *wie:UserAbility* characteristic should express this idea. This open world assumption must be anticipated through the explicitness of WIE modelling.

As the number of WIEs increases, so does the probability of having shared characteristics between them. To cope with this issue, the WIE meta-model introduces a specialisation mechanism (*WIEExtension*). For example, if two *WIE* instances share one or more characteristics, these can be refactored with the aid of *WIEExtension* instances, creating a new *WIE* instance encompassing the shared characteristics (and vice-versa). A more detailed explanation of working with *WIEExtension* is presented on Section 6, where a methodology helps teams exploiting the synergies between different WIEs.

A simple example of a *WIEModel* instance for a website is presented on Figure 7. Here, three *WIE* class instances are taken into account. First, the two main audiences, *Blind* and *Non-impaired* are represented with their specific characteristics. Afterwards, another *WIE* class instance is specified, *Generic User*, enclosing a set of shared characteristics between the two main audiences. The *WIEExtension* mechanism ensures this sharing property.

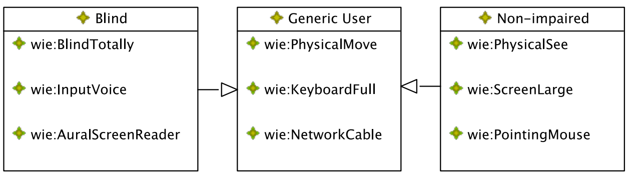

Figure 7: WIE modelling example

Next, we present a set of guidelines to model WIEs and to explore them for leveraging common grounds between WIEs.

6. GUIDELINES FOR WIE MODELLING

Having a characteristics vocabulary and an associated modelling framework by themselves merely provides a grammar to describe Web Interaction Environments. In order to define WIEs with proper semantics, i.e., with a useful *meaning*, we have formulated a set of guidelines to be followed. These allow WIEs to be grouped and abstracted into *common grounds*. Correspondingly, each common ground will represent the gist of the characteristics of its child WIEs. This has the consequence that common grounds will convey the shared *semantics* of its children, recursively. The exploration of these semantics by teams thrives the discussion of the several external factors that have impact towards providing the best universal user experience possible to everyone. Like Schneiderman has hypothesised in [27], *"accommodating a broader spectrum of usage situations forces researchers to consider a wide range of designs and often leads to innovations that benefit all users"*.

6.1 Defining WIE Classes

The first pass of this process is where teams decide which WIE class instances are to be supported. While the method of acquiring each one is specific to each scenario, classes should be specified as an instance of the *WIE* meta-class and attributed with a meaningful *name*. It should be noted that the number of instantiated WIE classes will influence both audience coverage and model complexity. The more classes defined, the more coverage will be obtained, but at the cost of increased complexity when deriving common grounds from WIEs. An example presenting class creation is depicted next, on Figure 8:

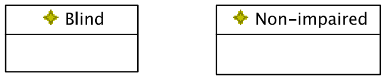

Figure 8: WIE class definition example

6.2 Selecting Characteristics

After all WIE classes have been defined, teams must select which characteristics fit on the concepts they have devised. This is done first by adding instances of the *WIECharacteristic* meta-class, whose *name* corresponds to a concept from the ontology presented earlier. It is worth noticing that the more characteristics a WIE encloses (either directly or indirectly through its ancestors), the more specific and complete is the user experience it affords.

The most important issue on selecting a characteristic from the ontology for a WIE class comes from the taxonomic property of the parent/child relationships between concepts (i.e., *is-a* relations). Based on this property, opposite selection strategies can be taken by teams, as follows:

- *Shallow selection*: going higher in the taxonomy (i.e., towards *Characteristic*, the root concept of the ontology), less expressive concepts can be selected. This leads to a compromise between coverage and specificity, where the higher a team goes, the more use cases can be covered by the enclosing WIE, but at the expense of being more loose regarding the unique aspects that characterise the environment (thus reflecting a less precise usability quality);
- *Deep selection*: contrasting with shallow selection, deep selection is performed by heading towards leaf concepts in the taxonomy, affording more expressive concepts to be selected. Accordingly, this also leads to a compromise between coverage and specificity. In this case, the deeper a team goes, less use cases are covered by the enclosing WIE. This might result in having to find additional WIE classes that cover selected characteristic's siblings on the taxonomy. However, the

expressed factor that characterises the environment reflects a more tailored user experience.

Based on these selection strategies, the second iteration of our example is presented on Figure 9, where several characteristics were added to each class, according to the selection criteria previously explained.

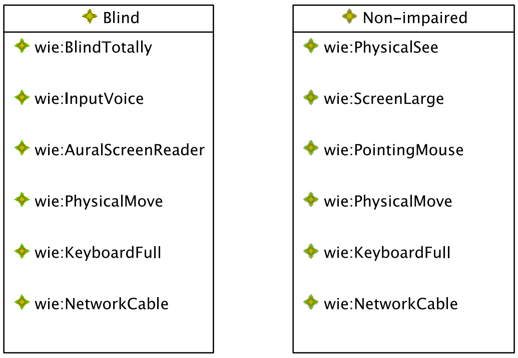

Figure 9: WIE characteristics addition example

6.3 Refactoring WIE Classes

The third pass of the process of establishing common grounds relates to extracting new knowledge from the concepts already specified on the first two passes (i.e., self contained and isolated WIE classes and their inherent characteristics). This is the most important pass of the process, since it affords the required expression of WIE modelling that allows teams to explore the synergies and differences between interaction environments. We have further decomposed this pass into three complementary strategies that leverage different aspects of the extension mechanism provided by the *WIEExtension* meta-class. These strategies can be applied recursively, and in no predetermined order. This should be approached exclusively depending on the scenario and use cases that are to be covered. The defined strategies are:

- *Replicas fusion*: when a set of characteristics is shared between a selection of two or more classes, teams should extract them into a parent WIE class. This must be done by creating a new instance of the *WIE* meta-class, the *parent*, and associate *WIEExtension* instances to the *children* (the mentioned selected classes) accordingly. Afterwards, the shared set of characteristics must be transposed into the newly created parent WIE class. This strategy is akin to well-known relational databases normalisation procedures [6];

- *Common grounds deduction*: this strategy merges *replicas fusion* with the *deep vs. shallow* characteristics selection earlier presented, through the exploitation of the *is-a* relationships of concepts on the vocabulary. Like *replicas fusion*, this strategy begins with the creation of a new instance of the *WIE* meta-class and associate *WIEExtension* instances between a *parent* and corresponding *children*. However, the criterion for selecting children does not reside on directly shared characteristics. Instead, teams should leverage an appropriate set of characteristics by applying a *shallow selection* on a cluster encompassing some or all of the children's characteristics, depending on which characteristics are interesting to be generalised. The corresponding *parent* WIE is the *common ground* between all of them.

  An important corollary for common grounds can be devised from the properties of characteristics selection criteria: since common grounds are *de facto* WIE classes, the selection of shallower characteristics tend to lead to a smaller WIE class count (thus reducing the WIE model complexity), but at the expense of being less expressive. Oppositely, deeper characteristics will narrow the scope of enclosing common grounds (thus affording more specific environments), but sacrificing WIE model simplicity. Therefore, teams are left with the best practice of choosing the appropriate coverage of common grounds.

Based on these steps, we have extended our example to fuse replicas and infer a common ground between the already defined classes. This common ground, as presented next on Figure 10, can be distinguished from the base assessment depicted earlier on Figure 7, by showing that a *wie:DeviceOutput* characteristic was deducted through a *shallow selection* strategy.

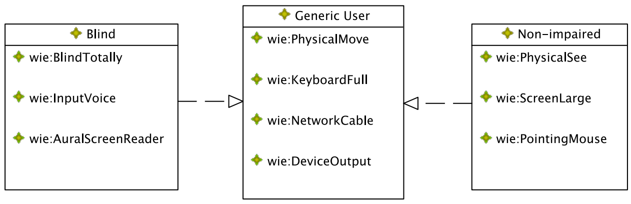

Figure 10: Refactored WIE model example

6.4 Leveraging Baselines

Finally, after classes have been defined and characterised, and after common grounds have been found, the *baseline WIEs* for a WIE model are leveraged. A baseline is defined as a special case of a common ground that represents the minimal of a set of WIEs that do not have a parent WIE class (either common grounds or isolated classes). In the case where only one WIE is found in this situation, it is automatically the baseline of the WIE model. While the process of finding the baseline is exactly the same to that of common grounds, an important distinction should be made. While common grounds are meant to represent synergies between WIEs at intermediate levels, baselines generalise what factors are to be taken into account in universal usability studies.

From this theoretical definition of a baseline WIE, one might devise a fundamental outcome: All WIEs can always have a baseline composed exclusively by the root concept from the vocabulary, *Characteristic*, by iteratively perform shallow selection operations on all WIEs until reaching this lowest common denominator. However, the information conveyed from this *root WIE* does not provide a meaningful answer. Therefore, we have defined that the convergence WIEs must be composed by a (non-strict) subset of the four characteristic domains' root concepts (i.e., *wie:User*, *wie:Device*, *wie:Situation*, and *wie:Intention*), in order to convey useful information.

Whether centred on root concepts or deeper ones, baselines leverage and make explicit two types of properties that characterise Web front-ends:

- *Extensiveness*: this property reflects the fact that the more characteristics a set of baselines has, the higher is the coverage of the universal user experience of a Web-front end. For instance, a WIE model encompassing a single baseline with the four characteristic domains' root concepts, implies that all WIE classes cover different characteristics in each domain. A common case of this situation can be verified on websites whose front-end is accessible to both non-impaired and blind users on desktop computers, as well as properly viewed on mobile phones;

- *Cohesion*: the other property that baselines make explicit relates to the number of baselines a WIE model has. If, by following the devised process, only one baseline is found, then the cohesion degree between all WIEs is maximal. This factor strongly impacts the usability of websites, as it implies a universality of front-end metaphors. On the other hand, when a website has more than one baseline, it indicates a split in universality, leading to the existence of profoundly different front-ends for each baseline. Such cases can also be verified elsewhere on the Web, where websites provide significantly different versions of front-ends to each particular audience (e.g., desktop *vs.* mobile).

Exploring the example that has accompanied this Section, we verify that the *Generic User* WIE is the minimal element of all WIEs, consequently representing the baseline WIE of the defined model.

7. CASE STUDY: WIKIPEDIA

We have devised an experience to study a particularly interesting subset of universal usability on the Web, its accessibility side. The following set of questions were formulated in order to understand the effect of allowing users to edit contents of webpages, including the addition of hyperlinks, whether point to internal webpages or to external ones:

- The average accessibility quality of Wikipedia webpages;

- The difference of quality between internal and external webpages;

- The probability of following a hyperlink whose quality is inferior to an initial webpage within Wikipedia;

- Which guidelines are more often followed in Wikipedia webpages.

7.1 Experiment

Using the modelling framework presented earlier, we have devised the WIE model representative of WCAG 1.0 checkpoints. This model is depicted in Figure 11.

Based in this model, we have devised a quantification process based on UWEM (Unified Web Evaluation Metholodogy) [33], encompassing a subset of checkpoints that can be fully automated (crucial for large scale studies). More specifically, the tests that were performed encompass the analysis of HTML structures, based on the guidelines provided in WCAG 1.0. This quantification provides answers to the audiences defined in the presented WIE model in a black-boxed manner. Consequently, the outcome results are representative of all audiences as a whole, not specific to a particular audience. The experiments were set-up in the following manner:

- A subset of 100 webpages from Wikipedia $S = \{s_1, ..., s_n\}$ was randomly selected;

- Each webpage p_i was crawled, in order to extract all of its the hyperlinks. Each hyperlink was followed, resulting on a set of webpages $S_i = \{p_1, ..., p_n\}$ for which s_i points to;

- A pre-processing task was applied to each webpage (more specifically JTidy[4]), in order to verify standard conformance. This process yields two metrics for each webpage x: E_x, representing the number of critical errors in the markup that could not be fixed, and W_x, representing the number of parsing warnings that could be safely ignored;

- All webpages were evaluated through a set of checkpoints $C = \{c_1, ..., c_n\}$ according to the following formula:

$$A = \frac{\sum c_i}{n}, c_i \in C$$

  where $c_i = 1$ when the checkpoint was successfully evaluated and $c_i = 0$ when the test failed. The result, A, yields a value in the range $[0, 1]$, representing a quality level for the each webpage.

7.2 Results

By choosing 100 randomly selected root webpages from Wikipedia (more specifically through a special hyperlink[5]), a total set of 7791 webpages were crawled. Out of these, 100 were randomly selected, and 7691 hyperlinks were followed. From the followed hyperlinks, 7211 were targeted to other webpages within Wikipedia, whereas 480 hyperlinks targeted to webpages outside the scope of Wikipedia. Table 1 synthesises the overall results for accessibility assessment:

	Number	Total	%
Webpages correct	772	7725	9.99
Checkpoints correct	5	14	35.71
Errors detected	66	7791	0.85
Warning detected	7782	7791	99.88

Table 1: Overall accessibility assessment

From these results, we verify that only nearly 10% of the webpages fully complied with the checkpoints. This situation would get definitely worse when applying more complex evaluation procedures. Out of the 14 evaluation checkpoints

[4] http://jtidy.sourceforge.net/
[5] http://en.wikipedia.org/wiki/Special:Random

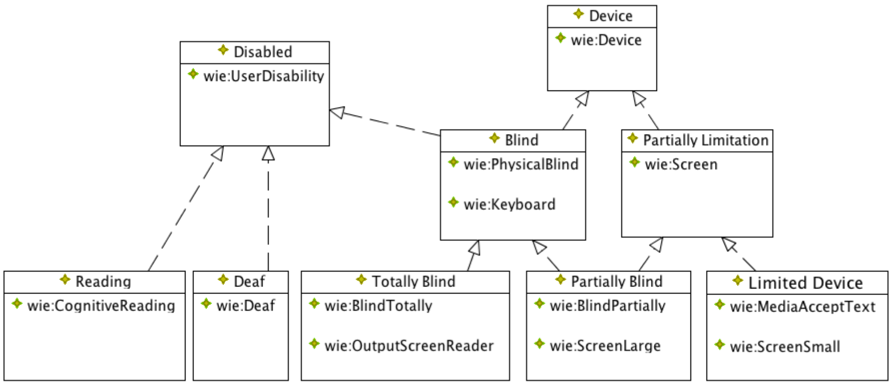

Figure 11: WIE model representative of WCAG 1.0 checkpoints

analysed, only 5 were fully complied by all webpages, based on UWEM tests: no *blink* elements, no *marquee* elements, no page refreshes through the *meta* element, the assurance that all *fieldset* elements have a *legend* child element, and the assurance that all *optgroup* elements have a *label* attribute.

The average quality of accessibility for each page was 84.6%, with $\sigma = 0.088$. The minimum accessibility quality obtained from the evaluated pages was 50% (i.e., 7 checkpoints), whereas the maximum was fully compliance.

When splitting the analysis between internal and external webpages, the results were different from the average values presented. Tables 2 and 3 further detail these findings:

	Number	Total	%
Webpages correct	750	7311	10.26
Checkpoints correct	5	14	35.71
Errors detected	0	7311	0.00
Warning detected	7311	7311	100.00

Table 2: Internal webpages accessibility assessment

	Number	Total	%
Webpages correct	22	414	5.31
Checkpoints correct	5	14	35.71
Errors detected	66	480	13.75
Warning detected	471	480	98.12

Table 3: External webpages accessibility assessment

Analysing both tables, it becomes relevant to emphasise the fact that while the number of checkpoints passed remains the same, the number of webpages which are fully compliant with all the 14 checkpoints evaluated is quite different between both clusters, with the ratio of 2:1 (internal:external). This is due to the fact that the overall HTML structure of Wikipedia complies with more checkpoints than the external webpages linked from it. Moreover, the fact that Wikipedia provides a simplified markup language, might provide additional support to these values. Also, as no parse errors were encountered in Wikipedia pages, a more standardised environment facilitates the usage of accessibility aids that require well-structured webpages. Based on these results, Table 4 presents a statistical comparison between both webpage clusters:

%	Internal	External
Average	89.79	81.83
Maximum	100.00	100.00
Minimum	57.14	50.00
Standard deviation	8.68	9.66

Table 4: Internal vs. external webpages accessibility assessment

Once again, the ratio of compliance between Wikipedia webpages and externally referenced webpages differs almost by 8%, and the minimum expected quality for any webpage within Wikipedia is more than 7% higher than externally references webpages.

Lastly, another analysis was performed based on clustering webpages between the initially 100 seed webpages and all the webpages linked from these. Tables 5 and 6 present their respective summaries:

	Number	Total	%
Webpages correct	13	100	13.00
Checkpoints correct	8	14	42.86
Errors detected	0	100	0.00
Warning detected	100	100	100.00

Table 5: Seed webpages accessibility assessment

	Number	Total	%
Webpages correct	759	7625	9.95
Checkpoints correct	5	14	35.71
Errors detected	66	7691	0.86
Warning detected	7682	7691	99.88

Table 6: Child webpages accessibility assessment

Both tables yield corollaries from the previous findings: if a random hyperlink is followed from a seed webpage, there is always the possibility of finding a webpage which cannot even be parsed (as child webpages include externally linked webpages). However, when comparing the statistics between both clusters, as presented in Table 7, more results can be studied.

These numbers help on reflecting about the exploratory nature of interacting with large content, highly linked web-

%	Seed	Child
Average	84.14	84.65
Maximum	100.00	100.00
Minimum	64.29	50.00
Standard deviation	9.27	8.76

Table 7: Seed vs. child webpages accessibility assessment

sites. If a user wants to follow a hyperlink from a Wikipedia webpage, there is no guarantee if the linked webpage has a higher or lower accessibility quality. However, it is mostly expected that, due to the wilderness nature of the Web vs. the highly structured and templated nature of Wikipedia, that there is a concrete probability that the minimum quality of accessibility will decrease significantly (15%, in the case of the experiment).

8. DISCUSSION

Based on the leveraged WIE model, as well as on the case study performed for Wikipedia, several issues should be discussed: the democratisation of Web authoring practices and its impact on Web accessibility and Universal Usability, the way quantification procedures work and how they can be entailed into large scale universal usability studies and, lastly, a reflection on whether guidelines or formal procedures improve universal usability on the Web in the long term.

8.1 Democratisation of Web Authoring

The immediate outcome of the experimentation performed over a random set of Wikipedia pages is two-fold:

- *Baseline quality*: by predefining easily usable template mechanisms, any user can markup content without having knowledge about Web accessibility best practices. Templates make sure that users introduce required contents, and template engines transform them in accordance to guidelines, which increases accessibility compliance;
- *Linking quality*: the other side of the democratisation of Web authoring concerns linking to external websites. Users are not aware that providing links to websites that have a decreased quality of accessibility results on a poorer experience from impaired audiences. Consequently, to mitigate such problems, these linking capabilities should make users aware of these problems, and suggest a set of alternatives and guidelines for providing external hyperlinks.

8.2 Quantification Procedures

Quantifying universal usability is not a trivial task. There is no current metric that affords measuring through quantitative manners the usability of a website for any audience, regardless of their particular characteristics. While several metrics have been proposed for interesting subsets of the whole spectrum of users (e.g., the visually impaired), there is still a large amount of work to be done. Moreover, as several procedures require manual inspection and verification of guidelines, quantifying universal usability becomes a daunting task when scaling up to the size of the Web. Hence, the limitations of these procedures also limit the knowledge that can be gathered and discovered about the Web.

8.3 Guidelines vs. Formalisation

It is a well known fact that current accessibility guidelines, such as WCAG, do not cover every aspect of accessibility, especially in what respects to audiences other than the visually-impaired [30]. Moreover, such guidelines only provide a baseline for usability, since both aspects contribute differently to universal usability [24].

When tying guidelines to quantification procedures, this becomes a critical issue. First, as explained in the case study, guidelines are typically black-boxed (they cover an entire spectrum of audiences within a particular topic). Consequently, every quantification procedure that builds upon them are also black-boxed, by definition. There is little knowledge about which checkpoint relates to which type of audience characteristic.

Consequently, there is an increasing need for the formalisation of guidelines that can be verified, assigned to specific WIE characteristics, and quantified automatically.

9. CONCLUSIONS AND FUTURE WORK

This paper presented *Web Interaction Environments* (WIE), a modelling framework to study universal usability on the Web. This framework encompasses a taxonomy of concepts that afford the characterisation of different audiences, as well as exploring their synergies and differences. Based on the WIE modelling framework, we have devised the model that remains implicit of WCAG 1.0. We applied a simple metric to quantify Web accessibility for the audiences represented by this model, in the context of a random selection of webpages from the Wikipedia website. This large scale study allowed us to verify that, despite being able to control several aspects of accessibility quality, template mechanisms such as those of Wikipedia cannot guarantee a high quality of user experience to the audiences covered by WCAG 1.0. Lastly, this paper has discussed a set of crucial points for enabling large scale studies of universal usability on the Web in general, and to Web accessibility in particular.

Based on the work presented in this paper, ongoing work is being done in the following fronts: (1) describe the semantics of relating different audience characteristics to evaluation checkpoints, based on their particular requirements; (2) define formal evaluation criteria for universal usability on the Web; (3) further explore other case studies in the context of universal usability and study the way quantification processes and guidelines differ in the context of different audiences; and (4) extend these studies to WCAG 2.0.

10. ACKNOWLEDGEMENTS

This work was supported by FCT, through scholarship SFRH/BD/29150/2006 and through the Multiannual Funding Programme.

11. REFERENCES

[1] R. Baecker, K. Booth, S. Jovicic, J. McGrenere, and G. Moore. Reducing the gap between what users know and what they need to know. In *CUU '00: Proceedings on the 2000 conference on Universal Usability*, pages 17–23, New York, NY, USA, 2000. ACM.

[2] T. Berners-Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt, and D. J. Weitzner. A Framework for Web Science. *Found. Trends Web Sci.*, 1(1):1–130, 2006.

[3] R. A. Botafogo, E. Rivlin, and B. Shneiderman. Structural analysis of hypertexts: identifying hierarchies and useful metrics. *ACM Trans. Inf. Syst.*, 10(2):142–180, 1992.

[4] B. Caldwell, W. Chisholm, J. Slatin, and G. Vanderheiden. Web Content Accessibility Guidelines 2.0. W3C Working Draft, World Wide Web Consortium (W3C), May 2007. `http://www.w3.org/TR/WCAG20/`.

[5] S. Ceri, F. Daniel, M. Matera, and F. M. Facca. Model-driven development of context-aware Web applications. *ACM Trans. Inter. Tech.*, 7(1):2, 2007.

[6] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, 1970.

[7] D. Dhyani, W. K. Ng, and S. S. Bhowmick. A survey of web metrics. *ACM Comput. Surv.*, 34(4):469–503, 2002.

[8] J. Fink, A. Kobsa, and A. Nill. Towards a user-adapted information environment on the Web. In *Multimedia and Standardization 98*, 1998.

[9] F. Garzotto, L. Mainetti, and P. Paolini. Hypermedia design, analysis, and evaluation issues. *Commun. ACM*, 38(8):74–86, 1995.

[10] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli. User Profiles for Personalized Information Access. pages 54–89. 2007.

[11] D. Heckmann. *Ubiquitous User Modeling*. PhD thesis, Department of Computer Science, Saarland University, November 2005.

[12] ISO/TS 16071:2003 - Ergonomics of human-system interaction – Guidance on accessibility for human-computer interfaces, 2003.

[13] M. Y. Ivory and M. A. Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.*, 33(4):470–516, 2001.

[14] H. R. K. Jahng, J.; Jain. Effective design of electronic commerce environments: a proposed theory of congruence and an illustration. *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, 30(4):456–471, 2000.

[15] G. Kappel, B. Prll, W. Retschitzegger, W. Schwinger, and T. Hofer. Modeling Ubiquitous Web Applications – A Comparison of Approaches. In *Int. Conf. on Information Integration and Web-based Applications and Services (iiWAS)*, Sep 2001.

[16] M. Kellar, C. Watters, and M. Shepherd. A Goal-based Classification of Web Information Tasks. *Proceedings 69th Annual Meeting of the American Society for Information Science and Technology 43*, Sep 2007.

[17] R. Lopes and L. Carriço. Leveraging Rich Accessible Documents on the Web. In *W4A'2007: the 4th International Cross-Disciplinary Conference on Web Accessibility*, Banff, Canada, 2007. ACM Press.

[18] L. Masinter, D. Wing, A. Mutz, and K. Holtman. Media Features for Display, Print, and Fax. RFC 2534, IEFT, Mar. 1999.

[19] J. Nielsen. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, 2000.

[20] Z. Obrenovic, J. Abascal, and D. Starcevic. Universal accessibility as a multimodal design issue. *Commun. ACM*, 50(5):83–88, 2007.

[21] Z. Obrenovic and D. Starcevic. Modeling Multimodal Human-Computer Interaction. *Computer*, 37(9):65–72, 2004.

[22] Z. Obrenovic, D. Starcevic, and B. Selic. A model-driven approach to content repurposing. *IEEE Multimedia*, 11(1):62–71, 2004.

[23] J. W. Palmer. Web Site Usability, Design, and Performance Metrics. *Info. Sys. Research*, 13(2):151–167, 2002.

[24] H. Petrie and O. Kheir. The relationship between accessibility and usability of websites. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 397–406, New York, NY, USA, 2007. ACM.

[25] J. Preece, H. Sharp, D. Benyon, S. Holland, and T. Carey. *Human Computer Interaction*. Addison-Wesley, Reading, Massachusetts, 1994.

[26] G. Schreiber and M. Dean. OWL Web Ontology Language Reference. W3C Recommendation, World Wide Web Consortium, February 2004. `http://www.w3.org/TR/owl-ref/`.

[27] B. Shneiderman. Universal Usability. *Commun. ACM*, 43(5):84–91, 2000.

[28] B. Shneiderman. Promoting universal usability with multi-layer interface design. In *CUU '03: Proceedings of the 2003 conference on Universal usability*, pages 1–8, New York, NY, USA, 2003. ACM.

[29] B. Shneiderman. Web Science: A Provocative Invitation to Computer Science. *Commun. ACM*, 50(6):25–27, 2007.

[30] D. Sloan, P. Gregor, M. Rowan, and P. Booth. Accessible accessibility. In *CUU '00: Proceedings on the 2000 conference on Universal Usability*, pages 96–101, New York, NY, USA, 2000. ACM.

[31] M. K. Smith, C. Welty, and D. L. McGuinness. OWL Web Ontology Language Guide. W3C Recommendation, World Wide Web Consortium, February 2004. `http://www.w3.org/TR/owl-guide/`.

[32] O. Troyer, W. Goedefroy, and R. Meersman. UR-WSDM: Adding User Requirements Granularity to Model Web-based Information Systems. In *First Workshop on Hypermedia Development, HT'98.*, 1998.

[33] E. Velleman, C. Meerveld, C. Strobbe, J. Koch, C. A. Velasco, M. Snaprud, and A. Nietzio. Unified Web Evaluation Methodology (UWEM 1.2), 2007.

[34] M. Vigo, M. Arrue, G. Brajnik, R. Lomuscio, and J. Abascal. Quantitative metrics for measuring web accessibility. In *W4A '07: Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A)*, pages 99–107, New York, NY, USA, 2007. ACM.

[35] M. Vigo, A. Kobsa, M. Arrue, and J. Abascal. User-tailored web accessibility evaluations. In *HT '07: Proceedings of the 18th conference on Hypertext and hypermedia*, pages 95–104. ACM, 2007.

[36] W3C Device Independence Group. Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0. W3C Recommendation, World Wide Web Consortium, January 2004. `http://www.w3.org/TR/CCPP-struct-vocab/`.