

Enhancing Wikipedia Editing with WAI-ARIA

Caterina Senette¹, Maria Claudia Buzzi¹, Marina Buzzi¹, Barbara Leporini²

¹ IIT – CNR, v. Moruzzi, 1, 56124 Pisa, Italy

² ISTI – CNR, v. Moruzzi, 1, 56124 Pisa, Italy

{Caterina.Senette, Claudia.Buzzi, Marina.Buzzi}@iit.cnr.it; Barbara.Leporini@isti.cnr.it

Abstract. Nowadays Web 2.0 applications allow anyone to create, share and edit on-line content, but accessibility and usability issues still exist. For instance, Wikipedia presents many difficulties for blind users, especially when they want to write or edit articles. In a previous stage of our study we proposed and discussed how to apply the W3C ARIA suite to simplify the Wikipedia editing page when interacting via screen reader. In this paper we present the results of a user test involving totally blind end-users as they interacted with both the original and the modified Wikipedia editing pages. Specifically, the purpose of the test was to compare the editing and formatting process for original and ARIA-implemented Wikipedia user interfaces, and to evaluate the improvements.

Keywords: WAI-ARIA, Wikipedia, user testing, accessibility, usability, blind users.

1 Introduction

In recent years wikis have enjoyed increasing popularity as collaboration tools in many areas. The most famous one is Wikipedia, the on-line Encyclopedia, which contains articles written collaboratively by volunteers from all around the world [22]. Building a single product in this way leads to a result with more integrated content [3]. In environments that allow many users to contribute to the creation of content, accessibility and usability are essential for universal participation. Wikis are a great opportunity for blind users, but the interactive environment and contents must be properly designed and delivered. Interacting with a wiki system can be difficult for a blind user since interaction requires the aid of assistive technology, adding a considerable degree of complexity.

We believe that an ARIA (Accessible Rich Internet Applications)-based editing interface would overcome many of the accessibility and usability problems that prevent blind users from actively contributing to Wikipedia. In previous phases of this study we discussed difficulties experienced by blind users when interacting with Wikipedia [5] and proposed applying the ARIA suite to enhance the usability of the editing page [4]. In this paper we compare the original Wikipedia Editing Page (WEP) and the proposed ARIA-based WEP in order to evaluate whether the latter simplifies editing and formatting. This comparison was carried out by means of a user test on both the editing pages involving totally blind end-users.

The paper is structured as follows: in Section 2 we present related works, in Section 3 we briefly summarize the problems of interacting via screen reader with the original WEP and describe changes to the HTML source file to improve accessibility/usability; Section 4 describes the user test environment and results. Lastly, conclusions and future work are presented.

2 Related Works

In recent years Wiki systems have become very popular as collaboration tools in several areas including eLearning, where Wikis are increasingly used as educational tools and/or integrated in Learning Management Systems (LMSs). However the effectiveness of Wiki-based systems as learning tools is still controversial: various studies have confirmed the effectiveness of Wikis in education [2], [3], [18] while others have revealed important limits [7], [8].

Soo-Hwan et al. [18] found that learning programming is more effective in the community setting than in a content-centered setting, and collaborative learning via Wikis is more valuable when learning requires problem-solving skills or the process of elaborating knowledge. By analyzing the use of two wikis in two Masters in IT, Bower et al. [3] verified that these collaborative tools can facilitate multi-user asynchronous creation, editing and restructuring of information To remedy the lack of interaction noted in online discussion groups, and to stimulate the collaborative environment, Augar et al. [2] adapted a traditional icebreaker exercise used in classrooms for use on a wiki, with good results: in the two-week exercise, the number of pages of the wiki increased steadily each day.

In contrast, other studies pointed out that wiki ideas and principles did not work in some learning environments, due to students' lack of motivation, no "community feeling", and poor usability [7], [8]. Specifically Ebner et al. [8] carried out two studies in higher education to investigate Wiki usage (student participation was voluntary, that is, users were not forced to utilize the system), highlighting "difficult to use" as one of the main causes of low active use of Wikis (inserting new content).

However, it is important to note that results of user tests directly depend on accessibility and usability of the UIs of the selected Wiki. The design and implementation of wiki-based learning tasks can greatly affect their success. Bower et al. [3] recommend a set of 12 principles for improving their effectiveness. For instance, these authors observed that tasks requiring a single product produce a wiki containing more integrated content. This is the case of Wikipedia, where each encyclopedia entry requires contributions by multiple authors, which are then integrated into a single description.

Usability is essential for the collaborative Web. Wikis' usability can be studied both by analyzing user navigation throughout pages (looking for the desired information) and by evaluating user interaction with the editing page (inserting/updating new content). In the latter case it was observed that in the initial life of wiki systems, an exponential growth of content starts only after an initial (and sometimes long) linear trend [7].

Wikis and e-Learning systems are a great opportunity for blind persons with mobility problems. However, electronic barriers can impede access to on-line resources. Accessibility guarantees that anybody can access Web content, regardless of any disability. This implies that different channels (visual, auditory, tactile) should be used to present the same content to the differently-abled.

The World Wide Web Consortium promotes accessibility on the Web through its Web Accessibility Initiative (WAI). Recently (Dec 2008) the WAI group produced a new version of the Web Content Accessibility Guidelines - WCAG 2.0 [23], which greatly improves the 1.0 version, and includes usability as a key factor, to be closely coupled with accessibility.

In order to verify Web page accessibility Takagi et al. suggest spending more time on the practical aspects of usability rather than focusing on the syntactic checking of Web pages. Indeed, some aspects are difficult to evaluate automatically, such as ease of understanding page structure and interface navigability [19]. To fill this gap, the WAI group is working on the Accessible Rich Internet Applications specification (WAI-ARIA) to make dynamic web content and applications (developed with Ajax, (X)HTML, JavaScript) more accessible to people with disabilities [24]. Using WAI-ARIA web designers can define roles to add semantic information to interface objects, mark regions of the page so users can move rapidly around the page via keyboard, etc. [24]. Useful examples of ARIA code are available on-line [25].

As previously mentioned, the design of any User Interface (UI) should include usability; a clear definition of usability is provided by the ISO 9241 standard [11]: “The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments”. There are many methods for evaluating usability, including heuristics, cognitive walkthroughs, guidelines, and usability testing.

In the study reported in this paper we describe a remote user testing conducted with a group of blind individuals in order to evaluate a modified Wikipedia editing page implemented with ARIA. Remote evaluation greatly reduces the cost of usability testing: in this type of usability evaluation, the evaluator performs all observation and analysis from a distance [10]. Different technologies for monitoring user behavior and capturing data can be applied, as shown in [17], [9] and [10], including videoconferencing, automatic logging of user paths and tasks, or others specific tools. Recent studies performing a comparative analysis have shown that during remote testing, users take a bit longer to complete tasks due to the communication overhead, but the results are as effective as, if not better than, traditional testing performed in the laboratory [21]. In contrast, Petrie et al. [16] conducted two case studies with disabled users (including totally blind and visually-impaired persons) to explore asynchronous remote evaluation techniques, and showed that while quantitative data were comparable, local evaluations collected richer qualitative data. However, these authors also argued that experienced specialists often lack a thorough understanding of how people with disabilities use their assistive technologies; thus perhaps the proposed questionnaire did not sufficiently reflect their needs and problems.

3. Modifying the Wikipedia Editing page (WEP)

3.1 Interaction via Screen Reader with the WEP

Blind users usually interact with computers (or other electronic devices) using assistive technology in the form of a screen reader. A screen reader is software that identifies and interprets the content being displayed on the screen, reproducing it through vocal synthesis (or rarely, by a Braille display). The screen reader reads a web page sequentially, one line at a time. When it works in “virtual mode” it is able to perceive that a browser is working, then tries to interpret the original Web page structure and give the user the navigation keyboard control. In this way blind users can halt the screen reading to scan a page with keyboard commands such as Tab key, from link to link, or via arrow keys to explore content line by line. In the following we refer to the JAWS screen reader with vocal synthesis since it is the most frequently used by the blind users in Italy [1].

The editing page of Wikipedia presents three main usability issues for totally blind users:

1. The formatting toolbar is difficult to perceive, access and use. The widgets of the Wikipedia toolbar are graphic icons, generated by JavaScript. The browser is unable to recognize these widgets as active elements (such as links, buttons, boxes, etc.) so they are skipped (never receive the focus) when the user explores the page via Tab key. Consequently, in this type of navigation the user never perceives the presence of a toolbar on the page. Instead, if the user explores the page sequentially (via arrow keys), the screen reader announces these widgets with the alternative description associated with the icons (for instance “graphic bold clickable”). However the user must be aware that each description/icon is associated with a formatting function and know how to activate it. This may be difficult for unskilled users. Another complex way to find a specific widget is to use the “find” command on the page. It is also possible to use other advanced screen reader commands but most users only utilize basic commands and are not even aware of this possibility [20].
2. It is difficult to select special characters and symbols. To insert a special character or symbol, Wikipedia offers a combo-box for selecting an alphabet. After the selection, a list of links of the corresponding alphabet characters is shown. Visually, this list is rapidly scanned, but since some alphabets contain more than one hundred links, it is not suitable for navigation via Tab key. Having many links makes the navigation long and users become disoriented. Furthermore, JAWS does not recognize uncommon symbols or characters, so it produces ambiguous text. For instance JAWS announces “link e” for each character in the group e, é, è, È and É. To distinguish each character, a more descriptive text should be associated (e.g. "e with acute accent") for enhancing usability.
3. Focus issue. In the WEP the focus is managed via JavaScript: when one or more words in the text area are selected, all related parameters (including the focus) are stored by the script in order to apply the formatting correctly. However,

when interacting via screen reader a blind user may not correctly understand how the focus is processed since the screen reader provides a “virtual focus”, and this may not coincide with the system focus. This problem could be quite important since to format a portion of text in the original WEP, the user must switch between editing and navigation modalities several times (see [4] for further details).

The screen reader is only one of the elements involved in the process of interaction between a blind user and a web page. In this process, blind people must understand the browser, the screen reader and finally the web page [20]. Visually-impaired people must make a great effort to obtain a good mental model of each of these three elements, especially if one of them is relatively unknown to the user. This problem is probably less relevant when they have to interact with desktop applications, but it becomes enormous when they navigate through the Web.

3.2 New Wikipedia Editing Page

To improve interface usability and solve the three issues described above, we introduced ARIA in the WEP UI source code. ARIA adds semantic information that communicates the object role (for instance role = “button”) to the screen reader and so to the user. In this way a graphic icon can be recognized as a control element. In this context the use of ARIA greatly enhances interface usability compared to using only standard (X)HTML elements (i.e. input element), making interaction via keyboard easier and more comfortable, as discussed in the following:

1. Formatting toolbar

The Wikipedia formatting toolbar, originally created with JavaScript, may be replaced with standard XHTML input elements with associated images (i.e. buttons), maintaining the graphical appearance of the original but providing accessible widgets. Access keys may be associated with each toolbar element to make it faster to apply formatting functions. However, memorizing 22 shortcuts (corresponding to the toolbar elements) costs the user significant cognitive effort, also because browsers and screen readers provide numerous shortcuts as well. Alternatively, several Tab key pressures are necessary in order to navigate the entire toolbar, so the UI is accessible, but navigation is still quite long and tedious. In order to simplify interaction, we have defined the formatting toolbar using the ARIA “toolbar” and “button” roles. The *activedescendant* attribute makes the toolbar navigable via arrow keys. Once the toolbar receives the focus via Tab key, the child elements -- i.e. each widget -- can be accessed by up and down arrows, and can be activated by pressing the ENTER key, which applies the associated formatting function (e.g. Bold, Italics, etc.).

2. Special characters and symbols

In the new UI we aggregate all characters of each alphabet in a second combobox located close to the alphabet combobox, as shown in Fig. 2 (right). Users first choose a ‘language’ and then select the desired character of that language.

This compact solution is faster for blind users since when navigating the combobox with arrow keys, the screen reader announces the character name

directly while in the original WEP every character read is preceded by the word 'link'. In addition when a user is in exploration modality, (s)he can skip/exit the combobox pressing a Tab key once, while in the original WEP it is necessary to cross/visit all the links. Lastly, the many links on a page make it difficult to use the special JAWS command Insert + F7 that gives the list of all the links in the page.

To simplify selection of a special character, we also specified a clear label attribute for each <option> item of the second combo-box, so the screen reader can announce a clear description of the selected character.

3. Focus

The focus problem is partially resolved by our new WEP. With the JAWS English version 9, the new interface allows users to insert and edit text without having to switch to navigation modality in order to find the active elements (widgets and comboboxes). The user activates the editing modality and it remains for the entire editing/formatting process, reducing the number of steps needed to complete the whole task. Instead with the JAWS Italian version 9 the focus is in the correct position but JAWS loses the editing modality. Indeed Jaws 10 automatically enables the editing mode when the virtual focus is in a text box (Auto Forms Mode) and a vocal tone is provided to the user to indicate the editing modality (Forms Mode on). Exiting the control element, the user automatically returns to the navigation modality. This makes interaction with form fields simpler and faster.

The ARIA *activedescendant* attribute (associated with the toolbar) allowed us to exploit the Wikipedia Javascript, which in the original UI was activated by the mouse click. Once the keypress event of a formatting button is captured, this Javascript applies the related function and moves the focus back to the text box.

4. User test

In order to evaluate the new Wikipedia Editing interface developed by applying our proposed solution, we conducted a user test with a group of blind people. Specifically, our evaluation aims to answer the following questions:

- Is the editing/formatting effectively simplified?
- Does the user perceive the editing/formatting task more quickly?
- Is the presentation of interface elements (combobox, labels) clearer?
- Which UI changes (i.e., ARIA toolbar or comboboxes) are more valuable for users?

Nielsen suggests that the following criteria are the most significant measures for evaluating a usability test [14]:

- performance (that is, the time required to perform a task)
- success rate (which measures whether users can perform the task at all)
- error rate
- users' subjective satisfaction.

In this study usability was analyzed based on the user's effectiveness and efficiency in carrying out a given task by blind users. Regarding performance, we

only measured the total time required by the user to complete each of two tasks on one of the WEPs (by using time-stamps server-side).

4.1 Design of the test

4.1.1 Method

To carry out the test we developed an environment that reproduced prototypes of both the original and the modified Wikipedia user interfaces. The environment also includes SW for recording and collecting useful information for the analysis of the user test results. The pages were developed in XHTML, PHP and Javascript. The system is mainly composed of the following components:

- A login page where the user can enter her/his name or nickname.
This (nick)name allowed all participants to be as anonymous as possible, which was greatly appreciated by the test subjects, who preferred not to be identified, and it allowed us to match together the test results with two questionnaires filled in by the users (described in the following).
The combination of recorded data with the users' subjective information (extracted by questionnaires) allowed us to better analyze the collected data.
The login page contained two buttons, one for the original and one for the modified UI, so the user could easily reach the two UIs (Fig. 1).



Fig. 1. Login Test Page

- Two Web pages reproducing the two WEPs – the original (Fig. 2) and the modified one (Fig. 3) – to be used for carrying out the tasks.
Since the main aim of our test was to compare user interaction with the editing field, the toolbar and the combobox, we had decided to remove additional text and links from the page, in order to focus users' attention on the features being evaluated. In this way we probably underestimate the advantage of ARIA that is particularly useful in complex UI.
For testing with Italian blind users, we used the Italian version of Wikipedia, downloaded on 2009 January (http://it.wikipedia.org/w/index.php?title=Wikipedia:Pagina_delle_prove&action=edit).

- A logging module to capture and record information on the test when the user navigated through the two WEPs.

The test produces two files for each participant: one concerning the interaction with the original WEP and one related to the modified one.

Figure 4 shows a snippet of the log file. Each time the user clicks on the "Save page", "Show Preview" or "End" buttons, the content of the text areas and a timestamp are memorized.

We added the "End" button to both the WEPs to force the user to register the server-side time before leaving the pages. This data is used to analyze how much time the user spent performing a task on a WEP.



Fig. 2. The original WEP



Fig. 3. The modified WEP

The test given to the blind users consisted of three parts, available in electronic format. The details about the test were sent to participants by email while tasks were executed via web, i.e. each user loaded from our server the original and modified UIs on their client via browser. In detail, the tests comprised three phases.

1. The first step of the test was sending the users a preliminary questionnaire.

This questionnaire aimed to collect information about participants, investigating their technical expertise, age, gender, computer environment (OS and screen reader), and use of Wikipedia.

Each participant had to choose a nickname for the identification.

Contextually we also sent user a URL for downloading and installing the Firefox 3.0.5 browser (being ready for the test).

2. The second step of the test was sending the users a document (Test Description Document) containing the test URL and directives for executing the test.

The test, executed by the user on his/her home computer (remote test), included two tasks to be performed on each WEP.

A task was considered finished when the user pressed the “save page” or “end” buttons. Data was appended in the file and elaborated after the test.

3. Last, user feedback was collected with a post-test questionnaire, consisting of 16 questions about:

- (1) difficulties in editing and formatting text
- (2) information about the subject’s experience performing the assigned tasks
- (3) the users' subjective satisfaction

The post-test questionnaire was sent to the users together with the document describing the remote test.

```
...
User name
...
Text area content
...
Text area end
Subject
...
IP address
82.57.143.XXX
Request time
25-01-2009 23:47:12
Browser
Mozilla/5.0 (...rv:1.9.0.5)... Firefox/3.0.5
...
```

Fig. 4. Snippet of the log file

Users were free to interact with the environment developed for the test (original and modified WEP) as they liked before the test, but we required them to perform each of the two parts of the test in one sitting (each part consisted of two tasks that were performed twice, on different WEPs). They could relax after completing each part in one of the WEPs and before starting the second part.

Users were divided into two groups (A and B) and each group began the test with a different WEP, in order to balance the average time spent on each WEP, because it is possible to assume that interaction with the second WEP assigned might be faster than the first. So we created two versions of the Test Description Document, and two of the post-test questionnaire, inverting the order of the WEP interaction. Since we did not know the number of the users participating in the test phase, we did the group assignment in progress, deciding on which group to assign a user only after (s)he sent back the preliminary questionnaire, and taking care to balance the groups.

4.1.2 Remote testing

We decided to set up a test where the tester and users were not co-located, i.e., remote testing as defined in [12]. According to Petrie et al. [16] classification, the remote testing was carried out asynchronously (i.e., the participant and the evaluator do not participate at the same time) guaranteeing participant independence (i.e., users undertake the evaluation independently).

The validity of remote testing, in comparison to classic laboratory usability testing, is a topic that frequently comes up in the literature. Both techniques have advantages and disadvantages, as discussed in [15] where the authors also suggest a “mixed” solution. We were forced to do remote testing since even a small change in the blind users’ hardware/software can dramatically impact on performance and thus on test results. In addition, with remote user testing we could easily recruit people from several Italian cities.

Furthermore, most of the subjects participating in the remote test reported that they did not feel comfortable and free to work if observed or monitored in any way during execution of the task. For this reason we choose a middle way, only recording times per task on each WEP. Data was created with server timestamps. It should be noted that performance measures require the homogeneity of the user hardware/software configuration and full control over user actions for test feasibility and validity of results, while qualitative testing is conceived for understanding user satisfaction and highlighting any problems. However, the test was based on the repetition by each user of the same task in both the UIs (the original and the modified) and on its evaluation.

This means that although executed in different environments (O.S. and screen reader version), the result is significant since the subjective improvement/worsening of any users in her/his interaction environment (user performance and satisfaction for the same task in the original and new UI) is evaluated.

4.2 The Test

To complete the test, each participant needed to connect to a URL (<http://test-utenti.iit.cnr.it/>) where they found a simple page with a login text box (password is not required) and the two buttons to the original and the modified WEPs (Fig. 1). The test consisted of two tasks to perform on both the UIs: (I) editing process using a special character; and (II) the use of a formatting function. Each task was split into a few steps to encourage the user to press a push button in order to allow us to register server-side a time-stamp in the logging file associated to each participant.

Tasks were balanced between users: half the people were asked to utilize first the original WEP and then the modified interface, and vice-versa for the other group. Table 1 contains the steps assigned to the users for the two tasks to perform in both the interfaces.

The right sequence of steps for each group was specified in the test instructions (Test Description Document). Without any temporal constraint, users were able to perform the test comfortably and without stress. Freedom to write any comment or

observation, in addition to the multiple-choice answers of our questionnaire, added very useful information to the analysis.

After preparing a preliminary version of the documents (Test Description Document and questionnaires) a pilot test set up by one of the authors of this paper, who has been totally blind since childhood, was performed. Other sighted authors also performed the test independently (with PC screen turned off), and afterwards outcomes were compared and discussed.

Thus, issues that might occur while performing the test (due to confusing descriptions/directives) were identified and fixed. The pilot test was an iterative process leading to the appropriate modifications to create an easier version of the test in relation to comprehension, set up and on-line access.

Table 1 shows test steps. Blind users had to write a simple article, the title was: “L’Italia” and the text was: “È una Repubblica” inserting the special character “È” from the combobox (task 1), and applying the formatting function Bold to the word: “Repubblica” (task 2).

We tested the interfaces with MS IE and Mozilla Firefox. We would favor the use of IE since test participants are familiar with this browser. Unfortunately, at the time of our test (January 2009) the new ARIA-based WEP was not correctly interpreted by IE v. 8 Beta while Mozilla Firefox v. 3.0.5 fully supports ARIA, so we required the use of the latter (Firefox). We tested the WEP with the Jaws screen reader v. 9.0. and v.10.

Table 1. Test description: steps of the tasks

T1 Editing task	
Step 1	Insert the word “L’Italia” in the Editing summary
Step 2	Select and insert in the Editing field the special character (È) using the combobox
Step 3	Continue to write the phrase: “ una Repubblica”
Step 4	Press the “Save page” button
T2 Formatting task	
Step 5	Apply bold formatting at the word Repubblica
Step 6	Press the “End” button

4.3 Participants

The user test was directly carried out with blind people, although it has been observed that working with sighted persons who have a certain expertise in computing and in using a screen reader could be a very effective method for testing accessibility problems [13]. In fact, anyone who regularly uses an assistive technology and perceives the content in a specific way can easily appreciate certain differences and features. However, we believe that a user interface developed for improving interaction by screen reader requires feedback from users who really interact with that assistive technology. In addition, the interface evaluation by usability experts may not be adequate, since blind users use the interface by responding to a different set of stimuli and criteria. For example, aural perception by a blind person is probably better than that of a sighted user, since evaluation of sounds and tones may be more

accurate. Therefore, in our opinion it is necessary to perform the evaluation with a group of blind users who actually interact with assistive technologies.

The participants were contacted through the Italian Association for the Blind, to which one of the authors also belongs. She initially contacted potential users via phone and email to describe the project and the aim of testing the new interface. In this way, 30 persons, completely blind since birth, took part in the first phase by filling out the preliminary questionnaire while 15 persons went on to participate in the entire test. In the following we will only refer to the 15 users who took the test.

4.4 Results

4.4.1 The preliminary questionnaire

Evaluating user feedback on a preliminary questionnaire is a basic procedure that can help to better understand the test results..

Data from the preliminary questionnaire provided a characterization of the sample. The 15 blind subjects comprised 5 women and 10 men, with age ranging from 18-24 years to more than 75 years (only 1 person) as shown in Figure 5.

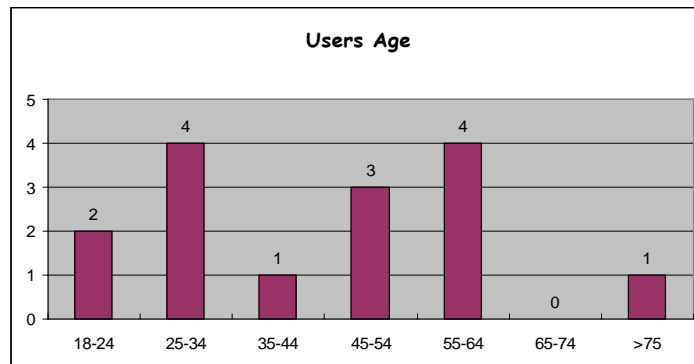


Fig. 5. Users' Age

Concerning the use of the Internet, 20% are beginners, 47% intermediate, and 33% had considerable experience. All the sample was using JAWS (from v. 6 to v. 9) on Windows (all use Windows XP; 3 of them also use Windows Vista).

In the test instructions we pointed out the need to use JAWS version 9.0 in order to benefit of the ARIA suite, but some people used a different version. For blind people in fact, it difficult to use a new version of assistive technology (different from their usual one), since most of them usually customize (with great difficulty) the screen reader according their specific needs or preferences.

It is obvious that different screen readers, as well as different versions of the same screen reader, may have different features, but their basic behaviors are similar. Furthermore, since evaluation of both the original and modified UIs was performed

by each user in the same environment, the results are comparable and the subjective evaluation has significance.

Regarding the use of JAWS advanced commands (for example: advanced command 'INSERT+F7' to have the link list, 'INSERT+F6' to have the heading list; 'N' to skip links), 40% of subjects use them always or frequently, 47% never or nearly never, 13% know only a few advanced commands.

Another of the questions included in the preliminary questionnaire was about the subjects' degree of knowledge of Wikipedia interfaces; answers highlighted that 20% of the sample don't know Wikipedia or know it only superficially, 40% are somewhat familiar with Wikipedia and 40% know it well. It is important to note that only four users (27% of the sample) have occasionally tried to use the Wikipedia editing page, which is the object of the test.

4.4.2 User Performance

We consider a test concluded if the user executed both the tasks regardless of correctness (e.g. applied the bold but inserted è instead of È). A total of 47% of the sample completed the test in the original WEP compared to 80% on the modified UI (Fig. 6).

A test was successful if the user concluded both tasks correctly in a WEP. The main observation is that 13% executed the test successfully in the original WEP compared to 33% on the modified UI (Fig. 6).

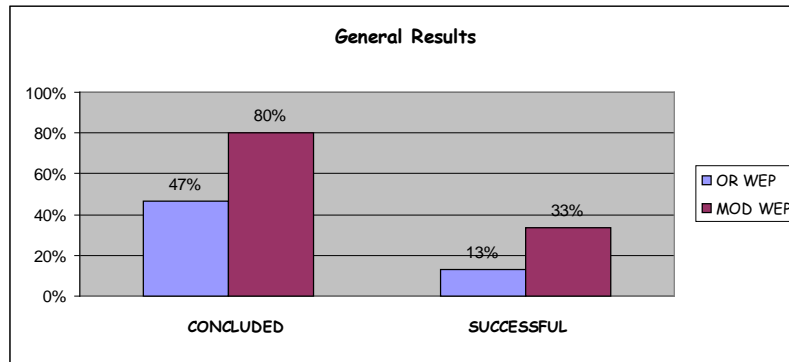


Fig. 6. Percentage of completed and successful tests

Figure 7 and Figure 8 show completeness and success of the two tasks.

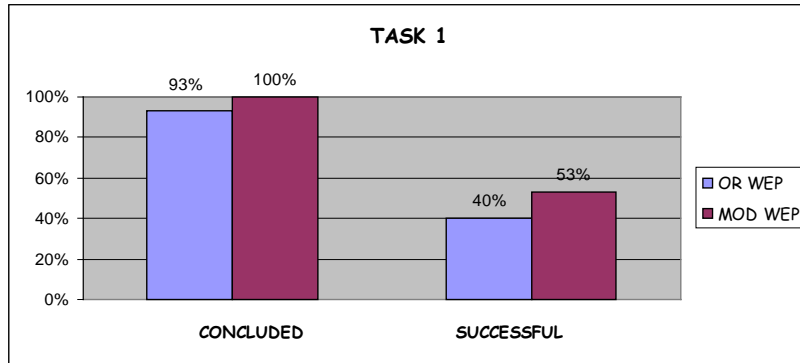


Fig. 7. Percentage of completed and successful task 1

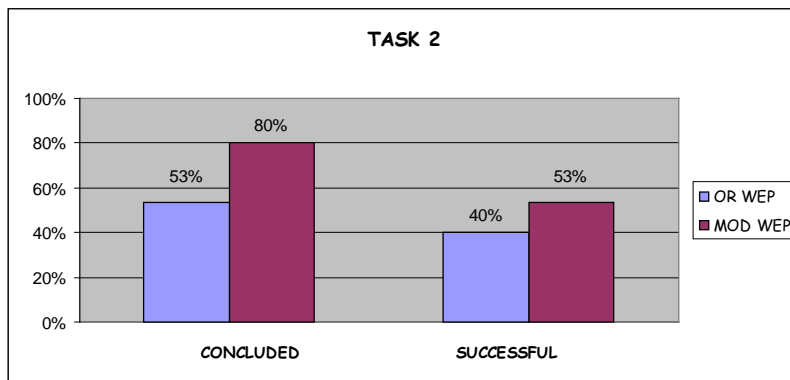


Fig. 8. Percentage of completed and successful task 2

These data indicate that the proposed UI appears to simplify interaction.

Regarding the execution times expressed in minutes, Figure 9 shows the average execution time for the entire test. Participants who completed the test reduced their execution time by 10% in the modified UIs.

One problem that most of participants faced was carrying out a test with an unknown UI. This means that part of the difficulty was experienced with the first UI used for the test (the original or the modified WEP), having to familiarize themselves with form elements. We believe that part of the long time it took to complete the first task (in both the original and the modified WEP) was due to the need to explore the UI. Another important factor is the use a different browser (Firefox) with respect to the usual navigation environment (IE).

It should be observed that in our test environment we only reported a small number (20) of all characters of the Latin alphabet actually present in Wikipedia (more than 150). For coherence of the test we restricted this number of characters to 20, and in the original WEP as well. This facilitates interaction of the original WEP compared to

the editing page of en.wikipedia.org. The latter presents more difficulties, thus the real advantage of the proposed combobox should be even greater.

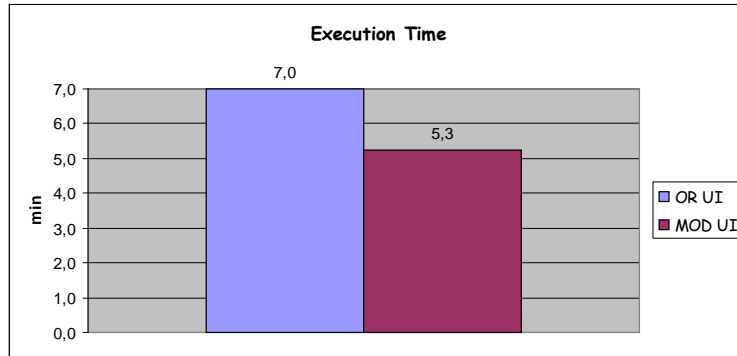


Fig. 9. Test execution time

Results in terms of user performance must be evaluated by looking at the data test and at the feedback that users gave us. To interpret results correctly it is necessary to understand when and how the participants performed the test, and also to consider all the factors involved, objective and psychological, when a person performs a test:

- 50% of the participants did the test at night; this may indicate that users needed a quieter environment but also implies that they may be tired or in a hurry;
- 80% of the sample did not experiment with the WEP before the test, possibly due to fear of interacting with the WEP, or to lack of time or interest.

Regarding the psychological aspect, which of course had great impact on performance, we can hypothesize following elements:

1. the user believes him/herself (and not the interface) to be the object of the examination. This explains why some users praise the modified interface although they have performed the tasks badly;
2. lack of self-confidence, probably due to inadequate feedback given to the user from the three software elements involved in the interaction: browser, screen reader and UIs (WEPs);
3. lack of naturalness while performing the tasks, since the user probably knows that someone is monitoring his/her behavior.

4.4.3 The Post-Test Questionnaire

The post-test questionnaire was formulated in a neutral way in order to avoid influencing the users, with the same question presented in two versions: one referred to the original and one to the modified WEP.

The post-test questionnaire reveals the user perception of the interaction with the WEPs and the user subject satisfaction. Of the sample, 60% found some differences while 40% found many differences between the original and modified WEPs.

Concerning difficulties with the test execution 73% found some difficulties, 20% many difficulties and only 7% no difficulties.

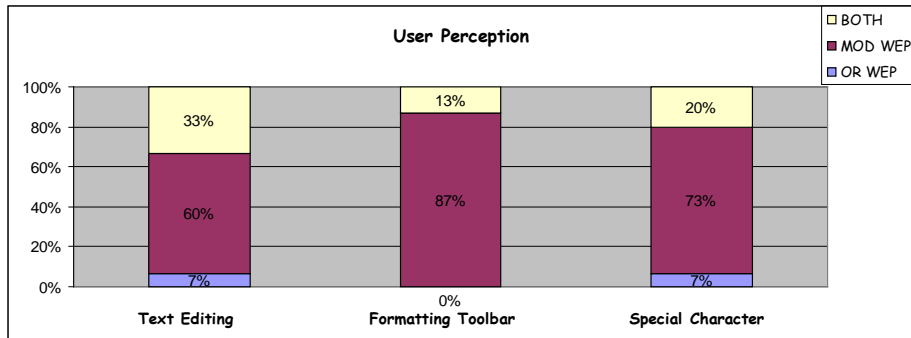


Fig. 10. User perception of rapidity of task execution

Some questions regard the user perception of rapidity of task execution: 1) editing text, 2) applying a formatting function 3) selecting a special character, as shown in Figure 10.

Those results revealed that most users appreciated the interaction with the modified UIs. Specifically, 60% of subjects declared that text entry was faster in the modified WEP, 33% in the original, and 7% found both interfaces equivalent.

A total of 87% of the participants declared that using the toolbar was faster in the modified WEP while 13% believed they were equivalent.

A total of 73% of subjects perceived the selection of a special character via combobox (instead of links) to be more efficient, 7% perceived the selection via links in the original WEP to be more efficient and 20% declared they were equivalent.

5. Discussion

In this paper we presented and discussed a usability test involving totally blind end-users regarding a proposed new Wikipedia editing page aimed at simplifying interaction via screen reader. The main purpose of our testing was comparing interaction via the screen reader JAWS with the original and new UIs developed using ARIA.

In this test we only evaluated two proposed changes in the original UI: an ARIA-based formatting toolbar and an additional combobox for selecting special characters.

Results of the test showed that interaction was simplified, since the number of completed tests is higher for the modified Wikipedia editing page. In the original editing page 47% of users concluded the test while the percentage increased to 80% in the modified one. Regarding correctness, 13% of the users concluded the test successfully in the original Wikipedia editing page while 33% of the sample did so in the modified interface. The interaction seems to be faster in the modified UIs, with a decrease in test execution time of 10%. Specifically, ARIA allowed more users to successfully complete the test and improved user performance.

However, the new UI would be expected to work much better than the original one, but results do not show great differences between the two WEPs. One problem is that

we are unable to verify whether some users did not follow our instructions closely, because we only measured users' final actions. In fact, we need to monitor all the events performed by the user (i.e. keyboard key or key combination press), in order to have real and total control of the experiment. Furthermore we forced many users to use an unknown browser (Mozilla Firefox) on a unknown UI (only 27% of users had previously used the Wikipedia editing page), with undoubted consequences for the results of the usability test.

User perception revealed that most users appreciated interacting with the modified WEP. Specifically 87% of the participants declared that using the toolbar is faster in the modified WEP while 13% declared they were equivalent. A total of 73% of the sample also perceived the selection of a special character via combobox to be more efficient than links.

6. Conclusions and future work

ARIA is an emerging technology that simplifies interaction for the differently-abled. Especially blind persons who interact via screen reader and voice synthesizer and the cognitively impaired may benefit to manage small portions of content anytime.

In the design of UIs all physical and cognitive resources required for the user interaction with all HW/SW components would be carefully analyzed. We believe it is crucial for designers to incorporate WAI-ARIA when developing user interfaces, to enhance usability via assistive technologies.

In this initial user test, we only worked on the main part of the Wikipedia Editing Page, removing additional text and links in order to focus users' attention on the features being evaluated. We have also used a restricted set of alphabets and characters, greatly simplifying the UI.

As a further step of this research we have completed the WEP with text, links, and special/alphabets characters, and we have added ARIA regions. In this new configuration, we anticipate considerable improvement in simplifying interaction with the Wikipedia editing page for the blind, since the current page layout is complex and crowded with active elements, making very useful for the blind users have a page overview (the list of the page regions) and be able to jump directly to the desired part of the UI (main contents), skipping navigation bar and menu (see [6] for a scenario of use)

However, at the moment we are unable to carry out a new test to evaluate our hypothesis, since JAWS v. 10, which supports ARIA regions, is not widespread among our test participants.

In the future, we intend to further utilize WAI-ARIA specification in other collaborative software as well, such as blogs and social networking sites.

References

1. Andronico P., Buzzi M., Castillo C., Leporini B.: Improving Search Engine Interfaces for Blind Users: a case study. Springer-Verlag Universal Access in the Information Society (UAIS), Vol. 5/1, pp. 23--41 (2006)
2. Augar, N., Raitman, R., Zhou, W.: Teaching and learning online with wikis. In: 21st ASCILITE Conference, pp. 95--104 (2004)
<http://ascilite.org.au/conferences/perth04/procs/pdf/augar.pdf>
3. Bower, M., Woo, K., Roberts, M., Watters, P.: Wiki Pedagogy - A Tale of Two Wikis. In: Information Technology Based Higher Education and Training Conference (ITHET06), pp. 191--202 (2006) DOI =10.1109/ITHET.2006.339764
4. Buzzi, M. C., Buzzi, M., Leporini, B., Senette, C.: Making Wikipedia Editing Easier for the Blind. In: ACM NordiCHI2008. ACM, pp. 423--426 (2008)
5. Buzzi, M., Leporini, B.: Is Wikipedia usable for the blind?. In: ACM W4A Workshop, ACM, pp. 15--22 (2008)
6. Buzzi, M., Leporini, B.: Editing Wikipedia content by screen reader: Easier interaction with the Accessible Rich Internet Applications suite. Informa healthcare. Disability and Rehabilitation: Assistive Technology, Volume 4 Issue 4, pp. 264--275 (2009)
7. Ebner, M., Zechner, J., Holzinger, A.: Why is Wikipedia so successful? Experiences in establishing the principles in Higher Education. In: International Conference on Knowledge Management (iKNOW), Knowledge Sharing in Research and Higher Education (KSR' 06), pp. 527--535 (2006)
8. Ebner, M., Kickmeier-Rust, M. D., Holzinger, A.: Utilizing Wiki-Systems in higher education classes: a chance for universal access? In: Springer Universal Access in the Information Society, Vol. 7, N. 4, pp. 199--207 (2008)
9. Hartson, H. R., Castillo, J. C., Kelso, J., Neale, W. C.: Remote evaluation: the network as an extension of the usability laboratory. In: ACM CHI 1996, ACM, pp.228--235 (1996)
10. Hartson, H. R., Castillo, J. C.: Remote evaluation for post-deployment usability improvement. In: Working conference on Advanced visual interfaces, pp. 22--29 (1998) [DOI = 10.1145/948496.948499]
11. International Organization for Standardization, Geneva: ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs), Part 11: Guidance on Usability, 1st ed., 1998-03-15 (1998)
12. Ivory, M. Y., Hearst, M. A.: The state of the art in automating usability evaluation of user interfaces. Computing Surveys, Vol. 2001, N.4, ACM, pp. 470--516 (2004)
13. Mankoff, J., Fait H., Tran T.: Is your Web page accessible? A comparative study of methods for assessing Web page accessibility for the blind. In: SIGCHI conference on Human factors in computing systems (2005) [doi>10.1145/1054972.1054979]
14. Nielsen, J.: Alertbox January 21, 2001. Usability Metrics. Available at <http://www.useit.com/alertbox/20010121.html> (2001)
15. Norman, K. L., Panizzi, E.: Levels of automation and user participation in usability testing. In: Interacting with Computers, Vol. 18, Issue 2, March 2006. Elsevier Science, pp. 246--264 (2006) [DOI = 10.1016/j.intcom.2005.06.002]
16. Petrie, H., Hamilton, F., King, N., Pavan, P.: Remote Usability Evaluations with Disabled People. In: SIGCHI conference on Human Factors in computing systems, pp. 1133--1141 (2006) [doi>10.1145/1124772.1124942]
17. Scholtz, J.: Adaptation of traditional usability testing methods for remote testing. In: HICSS 2001, IEEE, pp. 1--9 (2001)
18. Soo-Hwan, K., Hee-Seop, H., SunGwan H.: The Study on Effective Programming Learning Using Wiki Community Systems. Springer Lecture Notes in Computer Science, 4227, pp. 646--651 (2006)

19. Takagi, H., Asakawa, C., Fukuda, K., Maeda, J.: Accessibility designer: visualizing usability for the blind. In: 6th international ACM SIGACCESS conference on Computers and accessibility, pp. 177--184 (2004)
20. Theofanos, M. F., Redish, G.: Bridging the gap: between accessibility and usability. In: Interaction, vol. 10, Issue 6 (nov + Dec 2003), pp. 36--51 (2003)
21. Thompson, K. E., Rozanski, E. P., Haake, A. R.: Here, there, anywhere: remote usability testing that works. In: 5th conference on Information technology education table of contents, pp. 132--137, ACM Press, New York (2004)
22. Wikipedia. Wiki, <http://en.Wikipedia.org/wiki/Wiki>.
23. W3C. Web Content Accessibility Guidelines 2.0. <http://www.w3.org/TR/WCAG20/>, 5 Dec 2008.
24. W3C. WAI-ARIA Overview, <http://www.w3.org/WAI/intro/aria.php>
25. W3C. WAI-ARIA Best practices, <http://www.w3.org/TR/wai-aria-practices/>