Journal of Universal Computer Science, vol. 14, no. 9 (2008), 1513-1528
submitted: 30/4/07, accepted: 19/2/08, appeared: 1/5/08 © J.UCS

Testing Website Usability in Spanish-Speaking Academia through Heuristic Evaluation and Cognitive Walkthroughs

María Paula González
(National Council of Scientific and Technical Research CONICET, Argentina
VyGLab - Dept. of Computer Science and Eng. – Universidad Nacional del Sur, Bahía Blanca
Argentina
mpg@cs.uns.edu.ar)

Toni Granollers
(GRIHO Research Group - Universitat de Lleida, Spain
tonig@diei.udl.cat)

Afra Pascual
(GRIHO Research Group - Universitat de Lleida, Spain
afra@griho.net)

Jesus Lorés (in memoriam)

Abstract: Although usability evaluations have been focused on assessing different contexts of use, no proper specifications have been addressed towards the particular environment of academic websites in the Spanish-speaking context of use. Considering that this context involves hundreds of millions of potential users, the AIPO Association is running the UsabAIPO Project. The ultimate goal is to promote an adequate translation of international standards, methods and ideal values related to usability in order to adapt them to diverse Spanish-related contexts of use. This article presents the main statistical results coming from the Second and Third Stages of the UsabAIPO Project, where the UsabAIPO Heuristic method (based on Heuristic Evaluation techniques) and seven Cognitive Walkthroughs were performed over 69 university websites. The planning and execution of the UsabAIPO Heuristic method and the Cognitive Walkthroughs, the definition of two usability metrics, as well as the outline of the UsabAIPO Heuristic Management System prototype are also sketched.

Keywords: usability evaluation, UsabAIPO heuristic method, cognitive walkthrough, UsabAIPO metrics, Spanish-speaking context of use
Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

1 Introduction

Usability is a software attribute defined as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*" [ISO 98]. In this setting, a *context of use* is a description of the actual conditions under which the product (an interactive system) will be used in a normal situation, including cultural issues as those proposed in [Marcus 00; Shen 06]. Due to the growing relevance of usability as a key factor of

software quality, *usability evaluation* has proven to be crucial [Ivory 01]. Indeed, identifying the most general usability problems of a context of use can help not only to evaluate the interfaces belonging to it, but also to prevent usability errors when a novel interactive system in it is being developed. In the last years many efforts have been made towards focusing international standards, methods and ideal values related to usability to assess particular contexts of use, specially those involving webpages [Shen 06; Sturm 02]. However, to the best of UP partners' knowledge no proper specifications have been addressed towards the particular environment of the academic websites in the Spanish-speaking context of use[1] (SSCU, hereafter), which involves hundreds of millions of potential users (persons who speak Spanish either natively or by adoption) whose cultural background is primarily associated with the Spanish language and culture, regardless of ethnic and geographical differences.

Given the above situation, the *Asociación Interacción Persona Ordeandor* (AIPO)[2] is running the *UsabAIPO Project* (UP), a project centered on usability research which involves the participation of 15 multidisciplinary university research groups specialized in HCI, formed by Spanish scientists as well as researchers coming from other Spanish-speaking countries. The UP started in 2004 (for a presentation of results corresponding to the First Stage see [Lores 05]) and it is still underway. The UP ultimate goal is to promote an adequate translation of international standards, methods and ideal values related to usability in order to adapt them to SSCU, aiming at a later extension of results to other contexts of use related to web design in Spanish language. Currently, within the UP the context SSCU is represented by a sample group of systems formed by the web sites of the 69 universities listed in the Universia portal,[3] a widely used portal about universities available for Spanish-speaking countries. Besides, within all UP Stages four different categories were considered, namely *Design*, *Content*, *Navigation* and *Search*.

This article presents some statistical results associated with the Second and Third Stages of the UP, involving the application of the UsabAIPO Heuristic method (see Section 3) and seven Cognitive Walkthroughs (CWs) [Wharton 94] respectively. A sketch of some results presented here was previously published on [González 06]. Note also that for space reasons the qualitative processing of usability results is not considered here, although we are aware of its relevance (for more details, see [Gonzalez 07] and [Gonzalez 07b]).

This article is structured as follows. First, next section discusses some related work. Then, the Second Stage of the UP is presented, sketching the planning and execution of the UsabAIPO Heuristic method performed, as well as a discussion of some statistical results. A description of the metric *UsabAIPO-H* and an outline of the *UsabAIPO Heuristic Management System* prototype are also included. Next, the Third Stage of the UP is presented by discussing seven CWs defined for three different user profiles. The metric *UsabAIPO-CW* is sketched, as well as some statistical results. Finally, the paper concludes and discusses future work.

---

[1] The Spanish-speaking context of use is formed by Spain and all Spanish-speaking countries in America, including almost 35 million people living in the USA with Spanish-related cultural origins.

[2] See www.aipo.es (webpage of the Asociación Interacción Persona Ordenador AIPO).

[3] See www.universia.es

2 Related Work

In the last years many efforts were made towards evaluating usability in different contexts of use, focusing on premises coming from cross-cultural usability [Marcus, 2000]. Indeed, the notion of context of use was defined to bring together people with similar social and cultural behavior, even considering educational level and unemployment rates [Blažič 07]. In this setting, although Spanish is one of the most widely spoken languages in the world, only in the last years the definition of context of use for Spanish-speaking scenarios has emerged as a crucial field. To the best of the UP partners' knowledge, there is no similar approach to particularizing the existing international usability standards with a focus on SSCU as presented in this paper.

General recommendations on the basis of a particular cultural environment have been developed, especially those related to the question of which dimensions should be taken into account to localize a web page [Brandon 01; Bean 00]. Other interesting results are shown in [Shen 06]. This paper discusses the redesign of a computer interface for a Chinese user target group, incorporating a consistent and culturally rooted metaphor (a Chinese garden metaphor). As part of the experimentation a desktop interface (which is normally based on an American–English metaphor) and a web browser design are presented and ideal values for the context of use of the Chinese garden are outlined.

A case involving Spanish-speaking culture is presented in [Sturm 02], where an interesting comparison between homepages of three universities (the *Hogeschool van Amsterdam* in The Netherlands, the *Universidad de la República* in Uruguay and the *Instituto Tecnológico de Monterrey* in Mexico) is discussed. The screenshots of the three universities are compared in order to visualize localization related to the cultural level of internationalization. However, only a high-level analysis is presented, and no discussion about the ideal values for the corresponding contexts of use (i.e. homepages in Dutch- and Spanish-speaking academia) is included.

3 The UsabAIPO Heuristic Method

As pointed out before, the Second Stage of the UP consisted on the usability evaluation of SSCU through a *UsabAIPO Heuristic method* (UsabAIPO-HM), a particular usability evaluation method based on Heuristic Evaluation techniques [Nielsen 04]. Two general heuristic principles for each of the UP categories were defined and decomposed in several heuristic-related questions, all of them related to features to be evaluated. As a result, twenty-five questions were defined, five corresponding to the *Content* Category, seven corresponding to the *Design* Category, seven corresponding to the *Navigation* Category and six corresponding to the *Search* Category (see Table 1). Note that the proposal in Table 1 differs from the one shown in [Nielsen 04]. Instead of using traditional heuristic definition, partners of the UP decided to define the particular set shown in Table 1 (including ranges and ideal values for each possible heuristic-related answer) on the basis of 300 answers corresponding to an exhaustive poll carried out among users belonging to SSCU, and results coming from the UP First Stage [Lores 05] (in which different users and

scenarios in SSCU were studied and tested, and the usability of the homepages in SSCU were assessed).

Cat	*Usab-AIPO-HM Heuristic*	*Related questions*
D	Graphic Design	*Friendly Interface:* Does the website have a friendly interface, with uniform colors in most pages, matching with the university corporate image?
		Clean Interface: Does the website offer a clean interface, without visual noise and with a correct usage of space?
		Text Design: Does the text have a simple design, with enough contrast between background and text, limiting the font type and other text att.?
		Liquid Design: Is a liquid design being used?
	Images	*Labelled Images*: Are images labelled? Does their title appear when the mouse is moved over them?
		Animated elements: Do animated elements exist?
		Image Resolution: Has image resolution been taken into account, so that images are not pixeled and have an appropriate size for their correct visualization?
N	Naviga-tion Area	*Number of elements in menus*: Has the number of elements and terms per element been checked in order not to produce memory overload?
		Visibility: Is the totality of the elements in the navigation area visible without requiring any interaction from the user?
		Existence of site map: Is there a site map on the website?
		Affordance from links to applications: If a link leads to an application, is that clearly indicated?
	Orienta-tion	*Access to homepage*: Can the homepage always be reached from any navigation level?
		Orientation elements: Are there elements that allow the user to know exactly where he/she is in the website and how to move back? (breadcrumbs)
		Identifiable links: Do links clearly indicate where they lead to, using an appropriate title so that the user can predict the system response to his/her action?
C	Informa-tion	*News update*: Are academic news properly updated? Do they have publication date?
		Contact data: Is it easy to access to the information in the different areas of the university (secretary, departments, etc.)? Is there clear information about the data needed to contact some particular area (phone, email, etc)?
		News clarity: Are news items published on a salient place on the website, with a link to the news item associated with the news headline, and with a clear abstract of the contents of the news?
	Internatio-nalization	*Language:* Does the website offer a multilanguage option?
		Scope: The academic information available for the different languages is one page or is it the most part of the website?
B	Search Area	*Visibility and Simplicity*: Is it easy to start a search? Is the textbox used to search terms on the website to be found on the homepage? Is it easily accessible from any place of the website?
		Size: Is the text entry box for search wide enough? (it must contain between 15 and 30 visible characters)
		Complexity: Is there an advanced search option? Are there enough options to carry out a properly bounded search?
		Engines: Are links to Internet search engines included?
	Search Result	*Comprensibility:* Are search results shown in a clear form which is comprehensible enough for the user?
		Assistance: Is the user assisted when no results can be found?

Table 1: Heuristic-related questions for the UsabAIPO Heuristic method (D, N, C and S correspond to Design, Navigation, Content and Search Categories)

Then a specialized software tool called *UsabAIPO Heuristic Management System* (UHMS) was developed to support the usability evaluation of SSCU on the basis of UsabAIPO-HM (see screenshot in Figure 1). Indeed, the UHMS provides a simultaneous visualization of the current website that is being evaluated, the current heuristic-related question that needs to be answered (including possible answers), its significance within the website being considered and a pull-down menu to carry out the evaluation. To do this, every heuristic-related question was linked with an attribute in an output spreadsheet which stores the result of the evaluation. Besides, the tool provides a special form that allows the evaluator to add comments expressed in natural language, which will be automatically stored as part of the answer (see [Gonzalez 06]).

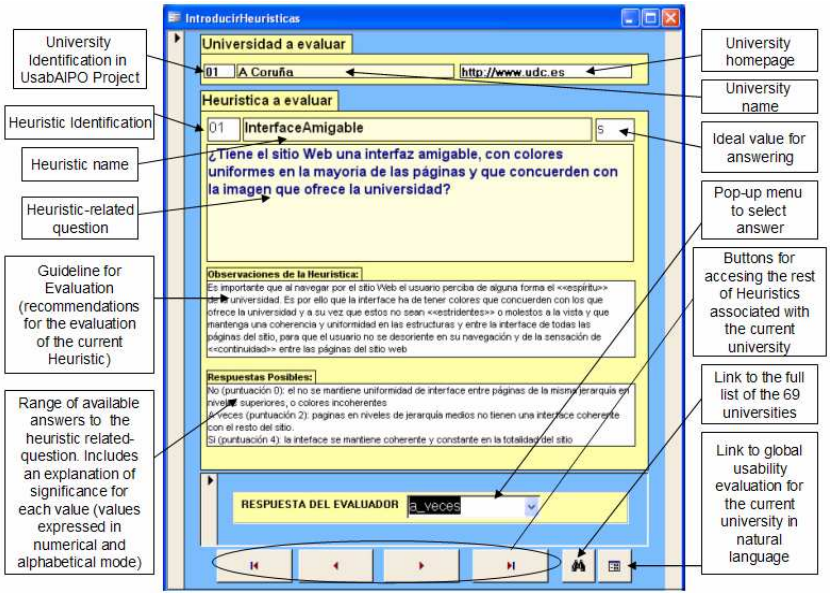

Figure 1: Interface of the UsabAIPO Heuristic Management System.

As a next step the UsabAIPO-H metric was defined to condense in a single value a general result for each system considered (each website belonging to SCCU). Note that values coming from each heuristic-related question must be normalized in order to compute UsabAIPO-H, as not all possible heuristic-related answers in UsabAIPO-HM are to be found between the same ranges. Moreover, each UP category was weighed according to the proportion of heuristic-related questions (the percentage of the total heuristic-related question performed, i.e. 25 questions), as follows:

$$UsabAIPO\text{-}H\ (w) = D*0{,}28 + N*0{,}28 + C*0{,}20 + S*0{,}24 \qquad (1)$$

In this equation w is a website belonging to SSCU, D represents the sum of the values obtained for the Design Category (28% of the total, i.e. 7*100/25=28%), N

represents the sum of the values obtained for the Navigation Category (28\% of the total, i.e. 7*100/25=28\%), *C* represents the sum of the values obtained for the Content Category (20% of the total, i.e. 5*100/25=20%), and *B* represents the sum of the values obtained for the Search Category (24% of the total, i.e. 6*100/25=24%). Next, the evaluation team carried out the usability evaluation of the 69 websites considered by applying UsabAIPO-HM. The evaluation team was formed by two usability experts (scientists with PhDs in the area of Usability Engineering) and two CS advanced students with solid knowledge about HE. All members were also frequent users of SSCU. The browsers MS Internet Explorer 6.0, Mozilla Firefox 5.0 and Netscape 8.0.1 were used to visualize the websites. The strategy used for each website evaluation included a first overview of the website under evaluation about 10 minutes, followed by an in-depth evaluation carried out by means of the UHMS. Two independent evaluations sub-teams carried out the 69 evaluations simultaneously. Results were compared and dicussed among the evaluation team members until a consensus was achieved (in some cases by looking for the median of the measured values). As a result, 1,725 data records were obtained (25 heuristic-related questions applied to assess 69 websites). The information was stored in a spreadsheet called *Usab_HE*, which was automatically generated by the UHMS. Then the UsabAIPO-H function was calculated on the basis of the data stored in *Usab_HE* (see Figure 2). According to UsabAIPO-H, only 2% of the websites in SSCU presented a usability value above 70 (ideal usability value: 100). Most websites (54%) scored around 60 (between 51 and 70), while a 13% of them reached values between 26 and 50. No website in SSCU was assigned a usability value below 26 or over 77.

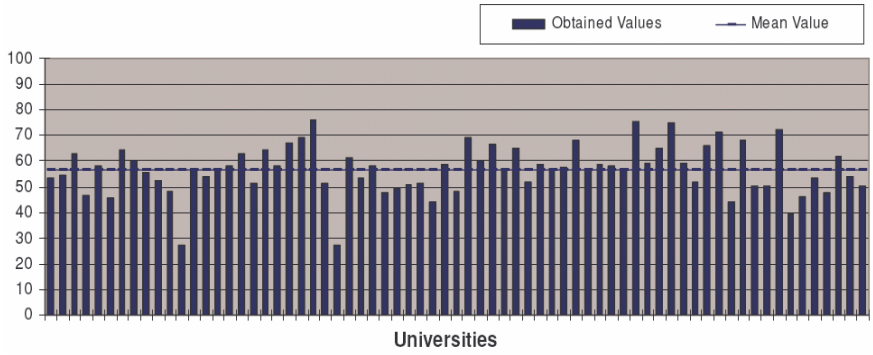

Figure 2: Results corresponding to the UsabAIPO-H calculation.

In order to study the UsabAIPO-H metric, a statistics analysis was performed over UsabAIPO-H results using the *Minitab* Platform.[4] The study included calculation of central tendency measures (mode, mean, median and quartiles), measures of dispersion (standard deviation, variance, obliquity and kurtosis), and a univariate analysis (considering distribution, the symmetry and presence of atypical values). Figures 3 and 4 depict some results corresponding to the calculation of central

[4] Licence No. 254052000082695, Universitat de Lleida.

tendency and dispersion measures. In addition, a multivariate analysis was included in order to study the relationship among the results corresponding to the different UP categories. In that respect, Figure 5 shows the general values obtained for each category (left) and particular values corresponding to the Content Category (right).

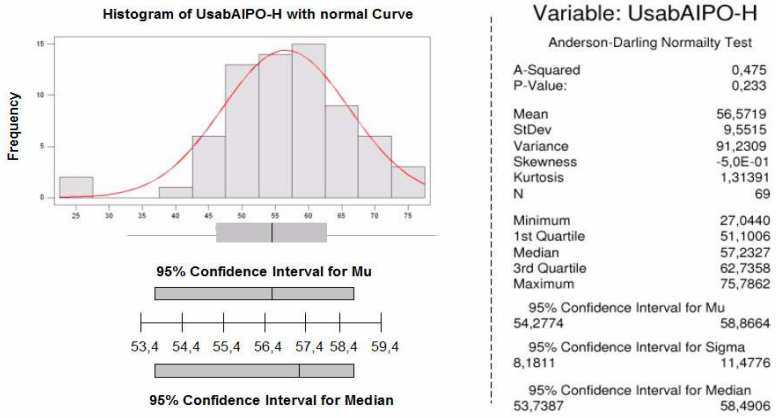

Figure 3: Statistical analysis of UsabAIPO-H. Left: Histogram and central tendency measures. Right: general results obtained.

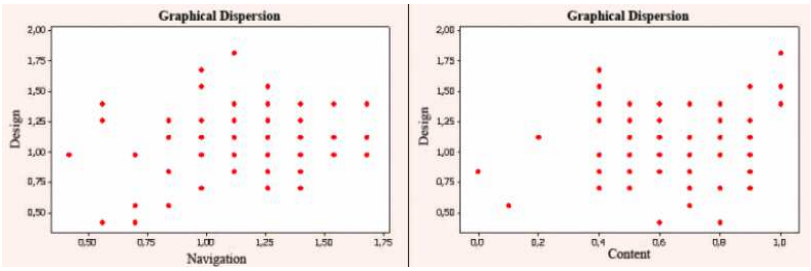

Figure 4: Statistical analysis of UsabAIPO-H. Some measures of dispersion.

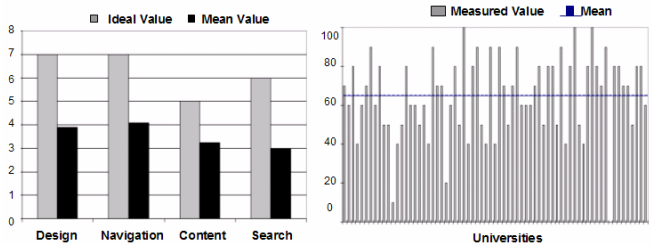

Figure 5: Multivariate Analysis of UsabAIPO-H. Left: general values for each Category. Right: particular values scored by the Content Category

Test Chi Square

Expected counts are printed below observed counts

	Design	Navigation	Content	Search	Total
1	4	4	11	13	32
	8,00	8,00	8,00	8,00	
2	14	9	22	27	72
	18,00	18,00	18,00	18,00	
3	25	23	25	22	95
	23,75	23,75	23,75	23,75	
4	26	33	11	7	77
	19,25	19,25	19,25	19,25	
Total	69	69	69	69	276

```
Chi-Sq =  2,000 +  2,000 +  1,125 +  3,125 +
          0,889 +  4,500 +  0,889 +  4,500 +
          0,066 +  0,024 +  0,066 +  0,129 +
          2,367 +  9,821 +  3,536 +  7,795 = 42,831
DF = 9, P-Value = 0,000
```

Figure 6: $\chi 2$ Test for statistical analysis of UsabAIPO-H.

Besides, the degree of independence of the categories according to the UsabAIPO-HM results was evaluated. The statistical function $\chi 2$ was used, calculated with 9 degrees of freedom (see Figure 6). Concerning the $\chi 2$ function, the independence of the four categories was proven, as it holds that (p-value)$<\alpha$ with α=0.05.

4 Cognitive Walkthroughs under the UsabAIPO Project

Cognitive Walkthroughs (CWs) [Wharton, 1994] involve evaluators inspecting a user interface by going through a set of tasks, assessing its understandability and ease of learning. Evaluators construct tasks and then play the part of a user working with that interface, "walking through" it. During the Third Stage of the UP, seven CWs were performed to assess the usability of SSCU. The evaluation team was formed by two usability experts and two Computer Science advanced students, all of them frequent users of SSCU. It must be noted that only one member of this team belonged also to the evaluation team described in Section 3. Even though the viewpoint of this particular member could be possibly biased by the previous evaluation, her participation was accepted as she accounts for only 25% of the new team and her general understanding about the UP was considered a good feature for the UP partners. Three user profiles were considered to adequate the definition of different CWs with respect to SSCU: *student*, *professor* and *administrative* profiles. Besides, seven CWs involving 37 tasks were defined on the basis of an exhaustive poll carried out during four months among 400 different users of SSCU. Eight questions were linked to each task (two questions for each category in UP). Note that for space reasons, only the definition of the seven CWs and the tasks related to the CW #1 are listed below. For a complete description of the 37 tasks see [Gonzalez 07c].

- CW#1 (student profile): localization and visualization of the study plan corresponding to a given undergraduate degree offered by the university under study./ CW#2 (student profile): localization and visualization of information concerning academic regulations (enrollment information, etc.)/ CW#3 (administrative profile): localization and visualization of information about a training course oriented towards the administrative staff and offered by the university/ CW#4 (administrative profile): localization and visualization of data for contacting a person belonging to the University administration using the option of People Search/ CW#5 (professor profile): localization and visualization of information about a posgraduate course or seminar offered by the university/ CW#6 (all profiles): localization and visualization of email access facilities/ CW#7 (all profiles): localization and visualization of a particular news made by the university.

- CW#1 (task description): *Task 1*: Visualization of types of degrees offered by the University (starting from the homepage); *Task 2*: Visualization of undergraduate degrees offered by the University (starting from the visualization of type of degree offered); *Task 3*: Visualization of study plan (starting from visualizing undergraduate degrees); *Task 4*: Short walkthrough using the study plan (courses, etc.).

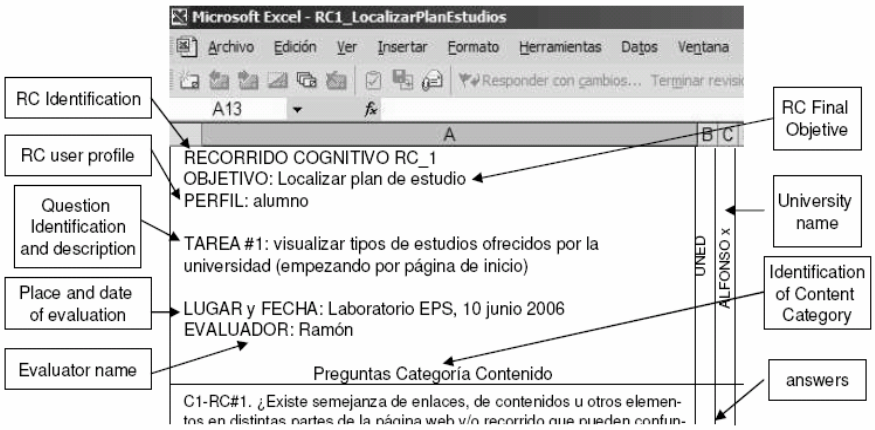

Figure 7: Zoom-in on the spreadsheet used to perform the Task 1 associated with CW#1.

Different ranked values for assessing users' cognitive effort were selected, namely (from bottom to top) *non_measurable, insignificant, low, normal, high and very_high* (corresponding numerical values form 0 to 5). Besides, ideal values between *insignificant* and *normal* were related to every task. Then, two auxiliary spreadsheets were designed to support the usability evaluation of the 69 selected websites on the basis of the 37 defined tasks. As an example, Figure 7 shows a zoom-in on the first questions of the Task 1 related to the CW#1. As in the UHMS, every

task-related question was linked with an attribute in these spreadsheets. Besides, the *UsabAIPO-CW* metric was defined to condense in a single value a general result for each website in C. Given a website *w*, the metric UsabAIPO-CW is simply the average of the seven values $E_1...E_7$, each of them representing the cognitive effort E_i assessed for a given CW#i when evaluating *w*. The cognitive effort E_i is calculated based on the values $T^i_1...T^i_{LastTask(i)}$, where LastTask(i) is a function that returns the number of tasks related to CW#i, and the values T^i_n are obtained when measuring each task n (with $1 \leq n \leq LastTask(i)$) during the performance of the CW#i. Note that each T^i_n is also weighted wrt a numeric value P^i_n to distinguish three different situations detected in the behavior of many users within SSCU: at the beginning of the CW#i, the users' cognitive effort increases, which seems to be due to their lack of familiarity with *w*; in the intermediate tasks of CW#i, this effort decreases as the users get accustomed to the interface of *w*; and in the last tasks of CW#i, the users' cognitive effort increases again (in this case maybe due to their mental tiredness associated with performing the task). Thus, the UsabAIPO-CW is defined as follows:

$$\begin{aligned} \text{UsabAIPO-CW}(w) &= (\textstyle\sum^{7}_{i=1} E_i(w))/7 \\ \text{with } E_i(w) &= T^i_1 * P^i_1 + ... + T^i_{LastTask(i)} * P^i_{LastTask(i)} \end{aligned} \quad (2)$$

where $P^i_1 > P^i_j < P^i_{LastTask(i)}$ for $j \leq LastTask(i)/2$, and $(\sum^{LastTask(i)}_{j=1} P^i_j)=1$. Next, the evaluation team was divided in two sub-teams in order to carry out two independent usability evaluations of the 69 websites considered by applying the redefined CW method described before. Due to the number of measures to be assessed, the execution of the seven CWs took about seven months. Again, the browsers MS Internet Explorer 6.0, Mozilla Firefox 5.0 and Netscape 8.0.1 were used to visualize the websites. The strategy used for carrying out each CW consisted in a first overview of the website under evaluation for about 10 minutes, followed by the performing of the CW itself. As a result, each sub-team collected 10,212 data records that were stored in temporary spreadsheets.

After processing all this data, each evaluation sub-team controlled the results produced by the other sub-team, and 24 data records were rejected as they had missing values. These missing values were related to unfinished CWs, which could not be completely assessed due to circumstances out of the scope of the evaluators (e.g. unknown password, nonexistent link, etc.). Afterwards both spreadsheets were condensed in a final spreadsheet called *Usab_RC*, containing 20,400 data records. The data were normalized in a scale from 0 to 100 to enhance the comparison of results corresponding to the Second and the Third Stage of the UP. Following this idea, also the function E_i had to be normalized by changing the condition $(\sum^{LastTask(i)}_{j=1} P^i_j)=1$ for $(\sum^{LastTask(i)}_{j=1} P^i_j)=100$. Then, the UsabAIPO-CW function was calculated on the basis of the normalized data stored in Usab_RC (see Figure 8). For the CWs involving three tasks, values to weigh the corresponding E_i function were selected as follows: 0.4 for weighting the first and third tasks, and 0.2 for weighting the second task. With respect to the CWs including four tasks, values to weigh the associated E_i function were chosen as follows: 0.375 to weighing the first and last tasks, and 0.125 for weighting the two intermediate tasks.

According to UsabAIPO-CW, only one university in SSCU presented a usability value above 70. Besides, 42% of the websites in SSCU scored between 60 and 70, while 50.56% of them reached values between 50 and 60. Besides, 6% of the cases had a value lower than 50; and no website was assigned a usability value below 35 or over 71. Note than for the UP Project, the ideal usability value associated with CWs after normalizing the data stored in Usab_RC should be between 30 and 40, as each of the seven CWs involved three or four tasks, each task ideally associated with a minimal cognitive effort (i.e. all $T^i_n=1$ for $1 \leq i \leq 7$ and $1 \leq n \leq$ LastTask(i), representing that all data related to the seven CWs were assessed with value=*insignificant*).

In order to study the above results, a statistics analysis was carried out over UsabAIPO-CW. This analysis included the same statistical study as the one calculated during the Second Stage of UP. As before, the *Minitab* Platform was used. Figure 9 depicts relevant results corresponding to the calculation of central tendency measures, including some numerical results associated with the measures of dispersion. Concerning the multivariate analysis of the obtained results, five functions were calculated for each of the seven CWs: a function to show general values obtained for each category, and four functions associated with values scored by each particular UP category (one category per function). As an example, Figure 10 shows general values obtained for each category for the CW#1 (left), and particular values scored by the Content Category for the same CW#1 (right).

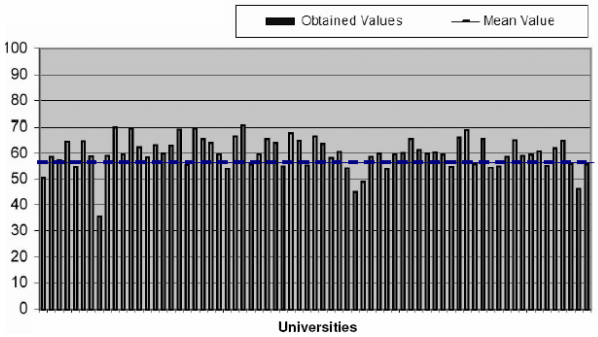

Figure 8: Results corresponding to the UsabAIPO-CW calculation.

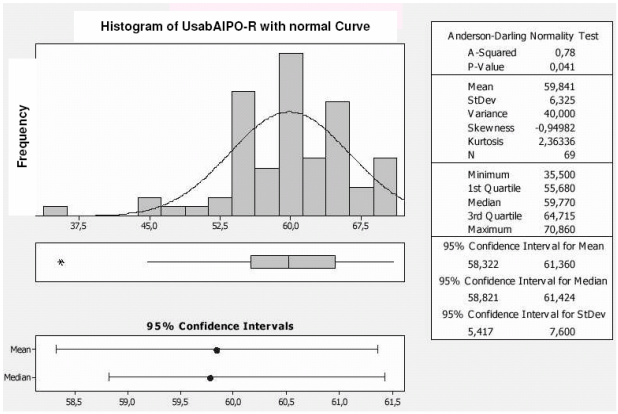

Figure 9: Statistical analysis of UsabAIPO-CW . Left: Histogram and central tendency measures. Right: general results obtained.

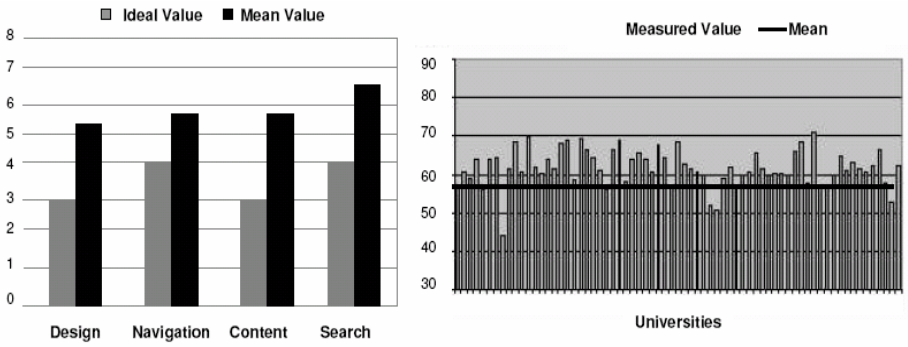

Figure 10: Multivariate Analysis of CW#1. Left: general values for each Category. Right: particular values scored by the Content Category.

Chi-Square Test

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	Content	Design	Nav.	Search	Total
1	14	25	10	27	76
	19,00	19,00	19,00	19,00	
	1,316	1,895	4,263	3,368	
2	14	22	19	17	72
	18,00	18,00	18,00	18,00	
	0,889	0,889	0,056	0,056	
3	28	15	28	7	78
	19,50	19,50	19,50	19,50	
	3,705	1,038	3,705	8,013	
4	13	7	12	18	50
	12,50	12,50	12,50	12,50	
	0,020	2,420	0,020	2,420	
Total	69	69	69	69	276

Chi-Sq = 34,073; DF = 9; P-Value = 0,000

Figure 11: $\chi 2$ Test for statistical analysis of UsabAIPO-H.

Moreover, Figure 11 shows results corresponding to the calculation of the $\chi 2$ function with 9 degrees of freedom. Again, note that the independence of the four categories was proven for UsabAIPO-CW, as it holds that (p-value)$<\alpha$ with α=0.05.

5 Conclusion and Future Work

Measuring usability of a given context of use as a whole is a challenging problem nowadays, and providing an appropriate solution for such a problem can be extremely helpful for software development and Usability Engineering in several respects. On the one hand, usability evaluators can rely on the usability features detected for a particular context of use as a whole when assessing a new interface belonging to it. On the other hand, the knowledge about usability features of this context of use can help software developers to prevent possible usability problems when a novel interactive system belonging to it is under development. Indeed, different contexts of use have been evaluated as a whole by specializing general usability methods, adapting international standards and ideal values to each particular environment under consideration [Shen 06; Sturm 02]. However, to the best of UP partners' knowledge no proper specifications have been addressed towards the particular environment of the academic websites in the Spanish-speaking context of use (SSCU), although this community includes hundreds of millions of potential users. To cope with this problem, the AIPO Association is running the UsabAIPO Project (UP), on which four categories (namely *Design*,*Content*, *Navigation* and *Search*) are considered. Indeed, the UP final goal is to promote an adequate translation of international standards and ideal values related to usability to adapt them to SSCU.

In this paper, some proposals included on the 2nd and 3rd Stages of the UP were described. A particular usability evaluation method called *UsabAIPO Heuristic* method (UsabAIPO-HM) based on Heuristic Evaluation techniques [Nielsen 04] was described and used to evaluate SSCU (currently represented in UP by a sample set of 69 websites belonging to the Universia Portal). Besides, seven Cognitive Walkthroughs (CWs) were defined and performed to assess the usability of SSCU.

Concerning the UsabAIPO-HM, we started by defining twenty-five usability-related questions to focus on SSCU, linking each question with one category in UP. Appropriate ranges and ideal values were established for possible answers on the basis of a poll carried out among 300 users of SSCU, plus previous research performed during the 1st Stage of UP (user testing and study of scenarios in SSCU). The specialized tool *UsabAIPO Heuristic Management System* was created to facilitate the UsabAIPO-HM execution, and the *UsabAIPO-H* metric was defined to condense in a single value the results assessed for each website in SSCU. Then, the UsabAIPO-HM was performed, 1,725 data records were collected and the UsabAIPO-H metric was calculated. A statistical study was carried out over UsabAIPO-H, including the calculation of central tendency and dispersion measures, as well as a $\chi 2$ test to prove the independence of the four categories considered. In the same way, 37 different tasks were defined to specialize the seven CWs. As before, usability related questions were associated with the UP categories, and an adequate range and ideal values were defined for each possible answer. Besides, the *UsabAIPO-CW* metric was outlined to summarize the results assessed for each CW, considering a final score for each evaluated website. Note that UsabAIPO-CW distinguished three situations related to different cognitive efforts that were detected in the behavior of many users within SSCU. After performing the seven CWs, the assessed values were condensed in a final spreadsheet called Usab_CW, containing 20,400 data records. Based on them, the UsabAIPO-CW metric was calculated. To study the obtained UsabAIPO-CW, a similar statistical study as the one performed during the 2nd Stage of UP was carried out. As before, the independence of the UP categories was proven.

Results coming from both 2nd and 3rd Stages of the UP showed the feasibility of applying both the specialized UsabAIPO-HM and the particular CWs. At the current stage of research, the obtained results account for a proof of concept of our proposal. However, more evaluation is needed to verify not only the obtained results, but also the correctness of the proposed specialized methods, metrics and ideal values. Indeed, part of our future work is focused on validating the performance of the defined UsabAIPO-HM and CWs, as well as the ideal values and usability-related questions here proposed. To achieve this, the 4th UP Stage is under consideration, involving the evaluation of academic websites of Spanish-speaking universities outside Spain under the same conditions established during the 2nd and 3rd Stages of UP. Certaintly, in the future 4th Stage of UP, the evaluation teams will be formed by members of SSCU belonging to countries ouside Spain, and will include real users corresponding to the differents profiles described during the 2nd and 3rd Stages of UP. Besides, the obtained results needed to be reported and discussed with different university webmasters in order to obtain their feedback and comments. In addtion, it must be remarked that results from 2nd and 3rd Stages of UP were just briefly contrasted, so that a deeper comparison is required to achieve more clear conclusions (including the analysis of

the inter-rater reliability of those people who are applying the proposed evaluation methods independently to assess SSCU, as well as a confrontation of the qualitative outcomes shown in [Gonzalez 07; Gonzalez 07b]).

Another research line currently explored is the deployment of a novel version of the *UsabAIPO Heuristic Management System* to be integrated in a *Usability Evaluation Management System*. Taking as starting point the existing interface, our goal is to develop an HE management tool oriented towards Spanish language which will include UP features by default and will allow us to easily redefine different elements, such as particular sets of heuristics, HE-related questions, and answers and alternative ranges and ideal values. Additionally, future work should explore the use of other usability evaluation methods (such as expert reviews based on personas and scenarios [Molich, 2007]) to assess SSCU. Comparing the results achieved by means of the 1st, 2nd, 3rd and 4th Stages of UP with this alternative data can also help to validate and enhance the presented proposal, especially if also its qualitative treatment (see [Gonzalez, 2007; Gonzalez, 2007b]) is included. In the same way, exploring the explicitation of culturally rooted metaphors for SSCU and its subsequent evaluation from a usability viewpoint would seem to be another interesting future research line to be explored.

Acknowledgments

This work was partially supported by CONICET (Argentina), Project PGI 34/N023 (Argentina); and Projects TIN2004-08000-C03-03 and SGR-00881 (Spain). We thank the collaboration of the different UP partners as well as of polled users.

References

[ISO 98]. ISO 9241-11. Ergonomic Req. Part 11: Guidance on Usability. (1998)

[Bean 00] Bean, J.: A framework for Globalization. Enterprise Development. (2000).

[Blažič 07] Blažič, B., Law, E., Arh, T.: An assessment of the usability of an Internet-based education system in a cross-cultural environment: The case of the Interreg crossborder program in Central Europe. J. Am. Soc. Inf. Sci. Technol., 58(1):66–75. (2007).

[Brandon 01] Brandon, D. (2001). Localization of web content. J. Comput. Small Coll., 17(2):345–358.

[González 06] González, M. P., Lorés, J. Pascual Almenara, A., Granollers, T.: Evaluación Heurística de Sitios Web Académicos Latinoamericanos dentro de la Iniciativa UsabAIPO. Proc. VII Int. Cong. Interacción Persona-Ordenador, ISBN 84-690-1613, pp. 145-157. (2006).

[González 07] González, M.P., Lorés, J., Granollers, T.: Assessing usability problems in latin-american academic webpages with cognitive walkthroughs and datamining techniques. In LNCS Springer Series. Proc. of the Int HCII'2007 (in press). Springer Verlag.

[Gonzalez 07b] González, M.P., Lorés, J., Granollers, T.: Enhancing usability testing through datamining techniques. Inform. Softw. Technol. (2007). (in press).

[Gonzalez 07c] González, M.P.: Integración de Técnicas de Descubrimiento de Conocimiento en Bases de Datos dentro de Modelos de Proceso de la Ingeniería de la Usabilidad. PhD Thesis. Copyright trasferred to Universitat de Lleida (in press, to be published electronically). (2007).

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

[Lorés 05] Lorés, J., González, M. P., Pascual Almenara, A.: First stage of the UsabAIPO project. Proc. VI Cong. Interacción Persona-Ordenador INTERACCION'05 (2005). (available in Spanish), pp 217–221.

[Marcus 00] Marcus, A., West, E.: Crosscurrents: Cultural Dimensions and Global Web-User Interface Design. Interactions, VII(4):32–46. (2000).

[Mayhew 99] Mayhew, D. J.: The Usability Engineering Lifecicle. A practioner's handbook for user interface desing. M. Kaufmann. (1999).

[Molich, 2007] Molich, R.; Dumas, J.S.: Comparative Usability Evaluation (CUE-4). Journal Behaviour & Information Tech. (2007) (in press)

[Nielsen 93] Nielsen, J.: Usability Engineering. M. Kaufman. (1993).

[Nielsen 94] Nielsen, J., Mack, R. (1994). Usability Inspection Methods. John Wiley & Sons.

[Rosson, 2002] Rosson, M., Carroll, J.: Usability Engineering: Scenario-based developement of HCI. M. Kaufmann. (2002).

[Shen 06] Shen, S., Woolley, M., Prior, S. (2006). Towards Culture-Centred Design. Interacting with Computers (in press), 19(2):1–33.

[Sturm 02] Sturm, C.: TLCC-Towards a framework for systematic and successful product internationalization. Proc. Int. IWISP'02. (2002).

[Wharton 94] Wharton, C., Rieman, J., Lewis, C., Polson, P.: Usability Inspection Methods, chapter The Cognitive Walkthrough Method: A Practitioners Guide, (1994). John Wiley & Sons Inc.