User experience methodology for the design and evaluation of interactive systems

Llúcia Masip Ardévol

Dipòsit Legal: L.855-2014
http://hdl.handle.net/10803/134762

Departament d'Informàtica i Enginyeria Industrial

Escola Politècnica Superior

User experience methodology for the design and evaluation of interactive systems

Thesis submitted in fulfillment of the requirements for the degree of Doctor in Computer Science by the University of Lleida

Llúcia Masip Ardévol

Supervisors of the PhD thesis:

PhD Toni Granollers Saltiveri and PhD Marta Oliva Solé

University of Lleida

December 2013

Universitat de Lleida

ACKNOWLEDGEMENTS

Although I decided to write the thesis in English, it is impossible to write the acknowledgements in a language other than Catalan since this is the language of my heart. However, there are people who helped me but they do not understand our language. So, the first part of the acknowledgement is in English and the rest is in the language that I am able to express everything sincerely, in Catalan.

Tot i que vaig decidir escriure la tesi en anglès, m'és impossible escriure els agraïments amb un altre idioma que no sigui el català, perquè així és com ho sento de veritat. Tot i així, hi ha persones que m'han ajudat durant aquests 4 anys però no tenen la sort d'entendre el nostre idioma. Per tant, la primera part dels agraïments està en anglès i la resta amb la llengua que sóc capaç d'expressar-me de cor, el català.

First of all, I would like to express my gratitude to the Interactive Critical System (ICS) Team from the Paul Sabatier University (Toulouse, France). Especially, I give my most sincere thanks to Marco Winckler and Philippe Palanque for giving me the chance of going to their research group. The three months there were a really positive experience that I will keep in mind forever. I met very nice people such as Regina, Martina, Camille, Celia, Michael, Eric, Caio, Arnaud, among others. They took me in as one more member of the group. *Merci beaucoup!*

Ara és el torn, de totes aquelles persones que en algun moment o altre d'aquests quatre anys també han fet possible que avui estigui escrivint aquestes línies.

A la Marta Oliva i al Toni Granollers (o al Toni i a la Marta, com sempre no sabia en quin ordre posar-vos) per donar-me l'oportunitat, ja fa uns anyets, de formar part del Grup de Recerca en Interacció Persona-Ordinador i Integració de Dades (GRIHO). Per voler assessorar-me i compartir l'experiència d'aquesta tesi. Per tots els moments viscuts tant bons com dolents, tant professionals com personals. Per tots, els consells que m'han donat i els dubtes existencials que m'han resolt sempre que ho he necessitat. Perquè els dos han estat imprescindibles. Mil gràcies.

Evidentment un agraïment molt especial a la resta de membres que han format o formen part del GRIHO: l'Anna Comas, Jonathan Giné, Juan Manuel Gimeno, Montse Sendín, Paco Jurado, Rosa Gil, Roberto García, Mireia Ribera, Esteve Brugulat, David Castellà, César A. Collazos i Andres Solano. Per tots els moments viscuts junts i per poder compartir aquesta experiència amb tots ells.

Durant la tesis he compartit unes quantes hores amb tres persones que han fet que aquests quatre anys hagin estat una mica més fàcils: l'Afra Pascual, companya indiscutible i confessora. El Josep Maria Brunetti també company tant pel dia a dia com pels congressos. I l'Anabel Usié, bona companya i millor amiga, no puc expressar tot el que hem compartit. Ànims si teniu la sort de no haver acabat i moltes gràcies per formar part d'aquesta tesi i de la meva vida.

També agrair al professor Josep Conde per la seva implicació en tots els còmputs estadístics. La seva aportació ha fet que aconseguís una de les parts més importants del projecte.

El Carlos Lozano, la Teresa Trepat i Josep Ramon Freixanet de l'empresa GFT van donar-me l'oportunitat i van tenir el coratge de voler provar l'eina i utilitzar-la en un entorn de treball real. Gràcies per provar-ho, pels comentaris aportats i per perdre una mica de temps i donar tant valor afegit a tot aquest projecte.

I ara és el torn de la família. Encara que siguin els últims no són, ni de bon tros, els menys importants. Les dos persones responsables de que jo hagi arribat fins aquí, d'inculcar-me els seus valors, de donar-m'ho tot sense demanar res a canvi, de donar-me l'oportunitat d'estudiar una carrera, un màster i donar-me el suport necessari per decidir-me i començar el doctorat són els meus pares, Jaume i Montse. Sense el vostre esforç això no hagués estat possible.

També agrair a l'Àngels per ser-hi sempre, cada dia i en qualsevol moment. En definitiva a tota la família, tan amb els que encara puc compartir moments com els que ja fa dies, o no tants, que ens van deixar. Tots sempre han estat convençuts que podia aconseguir-ho.

Finalment, i aquest sí que és per mi el més especial: gràcies Edu. Pel teu suport incondicional, per confiar amb mi més del que jo mateixa confio, per animar-me sempre i en qualsevol situació, per entendre el que faig i animar-me a millora-ho, per tenir clar que ho aconseguiria. Sense tu, impossible. Gràcies per ser-hi.

Simplement, gràcies.

SUMMARY

Heuristic evaluation is one of the most used and discount usability evaluation methodologies. However, it has some manual steps that could be semi-automated to decrease the time spent to carry out the methodology and, in consequence, the budget invested in the development process of an interactive system.

In addition, the quality of a product has evolved, in general terms, from the usability to the user experience. Therefore, according to both aspects, manual steps of the heuristic evaluation and evolution towards user experience, a new methodology based on heuristics is proposed for considering user experience in the design and evaluation steps of the development process of an interactive system.

This methodology manages all the needed information to perform the whole process of heuristic evaluation. An interactive system is defined through components, functionalities, features, user experience facets and attributes of the standard ISO/IEC 25010:2011. In addition, heuristics are related to components, functionalities, features, user experience facets and attributes of the standard ISO/IEC 25010:2011.

The methodology here proposed also semi-automates the selection of the best heuristics for a specific interactive system using the information commented before. If components, functionalities, features, user experience facets and attributes of the standard ISO/IEC 25010:2011 have heuristics and an interactive system is defined using components, functionalities, features, user experience facets and attributes of the standard ISO/IEC 25010:2011, the obtaining of the most appropriate heuristics for an interactive system is proposed.

Then, this new methodology supports the heuristic evaluation per se and, finally, it provides different UX measures. These measures get more reliable results through the best number of evaluators and more understandable results translating the wide range of results in four gravity levels. In addition, UX measures enable the standardization of the results for comparing the results of different versions of the same product or the same kind of interactive systems. Finally, it presents the quantification of the amount of user experience reached in a specific interactive system to know if the interactive system resulted in a positive experience for end users.

Furthermore, this methodology is implemented in Open HEuristic REsource for Designing and Evaluating User eXperience (Open-HEREDEUX). Open-HEREDEUX is

composed of four resources: Open Repository, Adviser of heuristics, Scorer of heuristics and Results Analyzer.

Open Repository collects all the needed information that Adviser uses to list the set of heuristics as appropriate as possible for designers or for evaluators. Then, if an evaluation should be carried out, Adviser sends the list of the most appropriate heuristics to Scorer. It will serve the evaluators for scoring all the heuristics. This module also saves all the information to be sent to Results Analyzer. Finally, Results Analyzer is the part of the system where results are processed through the *ISO/IEC 25062:2006 standard for software engineering. Software product Quality Requirements and Evaluation (SQuaRE). Common Industry Format (CIF) for usability test reports*.

RESUM

L'avaluació heurística és una de les metodologies d'avaluació de la usabilitat més barates i utilitzades. Tot i així, algunes de les fases de la metodologia poden ser semi-automatitzades per tal de disminuir encara més el temps necessari per executar la metodologia i, en conseqüència, disminuir també el pressupost invertit en el procés de desenvolupament del sistema interactiu.

A més, en termes generals, la qualitat d'un producte ha evolucionat de la usabilitat cap a l'experiència d'usuari. Per tant, d'acord amb les fases manuals de l'avaluació heurística i l'evolució cap a l'experiència d'usuari, es proposa una nova metodologia basada en heurístiques que considera l'experiència d'usuari en les fases de disseny i avaluació del procés de desenvolupament d'un sistema interactiu.

Aquesta metodologia gestiona tota la informació necessària per dur a terme una avaluació heurística completa. Un sistema interactiu es defineix mitjançant components, funcionalitats, característiques, facetes de l'experiència d'usuari i atributs de l'estàndard ISO/IEC 25010:2011. També semi-automatitza l'elecció de les millors heurístiques per un sistema interactiu específic utilitzant la informació presentada en el punt anterior. Si els components, les funcionalitats, les característiques, les facetes de l'experiència d'usuari i els atributs de l'estàndard ISO 25010:2011 contenen heurístiques i un sistema interactiu es defineix amb components, funcionalitats, característiques, facetes de l'experiència d'usuari i atributs de la ISO 25010:2011, es proposa l'obtenció de les heurístiques més adequades per un sistema interactiu concret.

A continuació, aquesta metodologia basada en heurístiques suporta el procés de l'avaluació heurística en sí i, finalment proporciona diferents mesures de l'experiència d'usuari. Aquestes mesures aconsegueixen resultats més fiables a través del número d'avaluadors més adequat i resultats més entenedors agrupant l'ampli rang de resultats en quatre nivells de gravetat. A més, les mesures de l'experiència d'usuari permeten l'estandarització dels resultats per poder comparar diferents versions del mateix producte o altres productes de caire similar. Finalment, és possible quantificar l'experiència d'usuari que s'aconsegueix en un sistema interactiu per tal de conèixer si el sistema proporciona, als usuaris que l'utilitzen, una experiència positiva.

Finalment, aquesta metodologia s'implementa en el recurs anomenat Open-HEREDEUX: Open HEurisitc Resource for Designin and Evaluation User eXperience.

Open-HEREDEUX està compost per quatre recursos: el Repositori Obert, el Recomanador d'heurístiques, el Puntuador i l'Analitzador de Resultats.

El Repositori Obert col·lecciona tot la informació que utilitza el Recomanador per llistar les heurístiques més adequades pels dissenyadors o avaluadors. A continuació, si es vol realitzar una avaluació, el Recomanador envia les heurístiques al Puntuador i, és aquest qui dóna suporta als avaluadors durant el procés de puntuació de les heurístiques. Aquest component guarda tota la informació de l'avaluació per tal de ser utilitzada en l'Analitzador de Resultats. Finalment, és l'Analitzador de Resultats qui obté resultats i els presenta seguint el model descrit a l'estàndard *ISO/IEC 25062:2006 standard for software engineering. Software product Quality Requirements and Evaluation (SQuaRE). Common Industry Format (CIF) for usability test reports*.

RESUMEN

La evaluación heurística es una de las metodologías de evaluación de la usabilidad más baratas y utilizadas. Sin embargo, algunas de las fases de la metodología pueden ser semi-automatizadas para disminuir aún más el tiempo necesario para ejecutar la metodología y, en consecuencia, disminuir también el presupuesto invertido en el proceso de desarrollo de un sistema interactivo.

Además, en términos generales, la calidad de un producto ha evolucionado de la usabilidad hacia la experiencia de usuario. Por lo tanto, de acuerdo con las fases manuales de la evaluación heurística y la evolución hacia la experiencia de usuario, se propone una nueva metodología basada en heurísticas que considera la experiencia de usuario en las fases de diseño y evaluación del proceso de desarrollo de un sistema interactivo.

Esta nueva metodología basada en heurísticas gestiona toda la información necesaria para llevar a cabo una evaluación heurística completa. Se define un sistema interactivo mediante componentes, funcionalidades, características, facetas de la experiencia de usuario y atributos del estándar ISO/IEC 25010:2011. También semi-automatiza la selección de las mejores heurísticas para un sistema interactivo específico utilizando la información presentada en el punto anterior. Si los componentes, funcionalidades, características, facetas de la experiencia de usuario y atributos de la ISO/IEC 25010:2011 contienen heurísticas y un sistema interactivo se define mediante los componentes, funcionalidades, características, facetas de la experiencia de usuario y atributos del estándar ISO/IEC 25010:2011, se propone la obtención de las heurísticas más adecuadas para un sistema concreto.

A continuación, esta metodología da soporte en el proceso de la evaluación heurística en sí y, finalmente, proporciona distintas medidas de la experiencia de usuario. Estas medidas consiguen resultados más fiables mediante el número de evaluadores más adecuado y también resultados más entendedores agrupando el amplio rango de resultados en cuatro niveles de gravedad. Además, las medidas permiten la estandarización de los resultados para que estos puedan ser comparados entre versiones de un mismo producto o entre productos similares. Finalmente, es posible cuantificar la experiencia de usuario que se consigue en un sistema interactivo con el fin de conocer si el sistema proporciona una experiencia positiva a los usuarios que lo utilizan.

Finalmente, esta metodología es implementada mediante el recurso Open-HEREDEUX: Open HEurisitc Resource for Designin and Evaluation User eXperience. Open-HEREDEUX está compuesto por cuatro recursos: el Repositorio Abierto, el Consejero heurístico, el Puntuador de heurísticas y el Analizador de Resultados.

El Repositorio Abierto colecciona toda la información que utiliza el Consejero para listar las heurísticas más adecuadas para los diseñadores y evaluadores. A continuación, si se pretende realizar una evaluación, el Consejero envía las heurísticas al Puntuador y, es este quien da soporte a los evaluadores durante todo el proceso de puntuación de las heurísticas. El Puntador almacena toda la información que será utilizada por el Analizador de Resultados. Finalmente, es el Analizador de Resultados quien los obtiene y los presenta siguiendo el modelo presentado en el estándar *ISO/IEC 25062:2006 standard for software engineering. Software product Quality Requirements and Evaluation (SQuaRE). Common Industry Format (CIF) for usability test reports*.

CONTENTS

FIGURES

TABLES

Chapter 1

"The shoemaker's son always goes barefoot"
"A cal sabater, el sabater és el més mal calçat"

Introduction

1.1 Motivation

My research career in the Human Computer Interaction (HCI) topic started in 2007 when I decided to study the Master's Degree in Human Computer Interaction at the University of Lleida. Only one year later, my current PhD supervisors (PhD Marta Oliva and PhD Toni Granollers) considered me to work in GRIHO (HCI and Data integration Research Group from the University of Lleida). I worked on the annual collaboration between the Lleida City Council and the University of Lleida, among other technology transfer projects. The Lleida City Council wanted to evaluate the usability of two new public interfaces that offered services to their citizens. At this moment I started a real project in the HCI topic.

The first experience concerned the usability analysis of a virtual website assistant. The second referred to the usability and accessibility analysis of interactive physical devices called Citizen Information Points (CIPs). The previous processes that I mentioned were the real motivation for starting my thesis.

Therefore, according to our knowledge in HCI and the budget provided by the collaboration, we decided to use the heuristic evaluation methodology as usability discount methodology to evaluate interactive systems in a quick and cheap way.

In the next section the real case studies proposed by the Lleida City Council are detailed. Both experiences will be cited during the whole document.

1.1.1 Berta: the virtual assistant

Berta is a virtual assistant provided by the Lleida City Council in its website (http://www.paeria.es). Its main function is to help users to find information about

different online procedures that citizens can carry out in Lleida using its website. Figure 1.1 shows the virtual assistant interface.

The Computer Department of the Lleida City Council hired us because they were worried about the usability level of their virtual assistant. After starting, those responsible warned us that they can change the assistant's face and the dialogs used but they cannot modify the interface codification, since it is an external application.

Thus, they were not interested in usability problems concerning the interface design. Their main interest was about facial expression problems and dialog problems. Take into account that facial expressions are connected with the dialog used and all this depends on the vocabulary typed for end users.

With this information, we studied their needs and we were aware that we could not use the same heuristics that we use for common websites. We had a different interactive system with different features and our usability goal was completely different from previous experiences. It was an excellent challenge for us.

Figure 1.1 The virtual assistant called Berta

Finally, we decided to perform a heuristic evaluation rejecting common heuristics from Nielsen [NIE94]. We did not consider it because it did not cover the main objectives of our usability evaluation and we believed that the most popular heuristics [NIE94] were not suitable for corporal expressions, dialog and vocabulary.

1.1.2 Citizen Information Points

The second real experience was the usability and accessibility evaluation of interactive public kiosks called Citizen Information Points (CIPs). CIP is a physical device located in public buildings spread throughout the city. Figure 1.2 shows one of the CIP devices.

Every CIP enables carrying out various public procedures, such as obtaining certificates or reviewing personal and local information. CIPs are also provided with a printer, keyboard, mouse, screen and digital signature reader through cryptographic card, electronic identity card or USB flash drive. It is like a cash machine, but with computer capabilities and redesigned elements for those specific locations and situations.

Figure 1.2 Citizen Information Point in its location

In this case, the main goal was not the screen interface but to evaluate usability and accessibility characteristics according to their specific localization and their particular physical features.

Thus, we detected that we could not use the most common heuristics [NIE94] because these heuristics, in the same way as the experience above, did not cover our evaluation goals and all the features of CIPs.

In addition another important challenge was to prepare how the heuristic scoring would run. As we commented, CIPs are spread throughout the city and the evaluation should be done in situ. So, the evaluation per se was done with a paper-based process.

1.1.3 Real experience discussion

As we have seen, every interactive system has its own features and usability goals that widely differ between common websites and virtual assistant interfaces or physical devices. The first aspect we noticed was that we could not use the most common heuristics (like Nielsen's) [NIE94] in both experiences because these heuristics did not cover our evaluation aims and all the Berta and CIP features. Furthermore, the bibliography studied did not provide us specific rules for guiding our studies. Then, our first difficulty was how awkward it was to choose the most suitable heuristics for those cases. As far as I know, there is no formal and specific definition about usability heuristics for these types of interactive systems, and that was the first research line.

Obviously, in both cases, the extraction of results was completely manual or using the basic functions of a spreadsheet program.

Moreover, some feelings flew in my mind during the execution of both projects. I was working on the improvement of the interaction of some interfaces but I was not using any type of interface that helps me in this process. We live in the 21th century where the technology revolution is underway but I was carrying out the evaluation using paper and pen. How was it possible that we (as a community that improves the life of users facilitating interaction with some types of devices) do not have any technological resource to carry out this kind of project? And then, I remembered the popular saying: "*The shoemaker's son always goes barefoot*". Thus, I acquired my PhD topic.

Finally, as the conclusion of this section, it is important to remark that both experiences are used as case studies and test cases for different parts of my PhD due to its different nature compared to the website and desktop applications. Therefore, references to these cases will appear in the entire document.

1.2 Objectives

The general goal of this research is:

> The definition of a quick and economical methodology which permits the consideration of the user experience in different steps of the development process of an interactive system.

Based on one of the most used methodologies [UPA09] to evaluate the usability of an interactive system, the heuristic evaluation, the adaptation and the optimization of this methodology are proposed.

Specifically, the main goal has two sub-goals to be achieved:

- The semi-automation of some parts of the heuristic evaluation methodology. Although heuristic evaluation is considered as a discount methodology to evaluate the usability of different interactive systems, some steps of the methodology are completely manual. Therefore, if the automation of these steps is possible, the heuristic evaluation methodology will become a faster and more economical method and as a result, a more cost-effective methodology.

- Nowadays, when designing and/or evaluating interactive systems we do it in terms of User eXperience (UX) instead of only usability. Then, the adaptation of the heuristic evaluation methodology for the consideration of more features besides the usability is also proposed.

Therefore, the main contributions given in the next section detail how the adaptation and optimization are reached.

1.3 Main contributions

Bearing in mind the goals of the research: the automation and adaptation of the heuristic evaluation methodology, the main contributions are:

- **The *definition of a methodology to consider the user experience in any step of the development process of an interactive system***. Most usability methodologies are often used either in the evaluation step or in the design step, and it is very difficult to apply or adapt the use of the methodology in other steps of the development process. Up to now, our methodology could be applied in the design and evaluation step and it is easy to adapt to other stages such as in the prototyping phase. (Chapter 3)

- **The *adaptation of the heuristic evaluation to consider more concepts apart from the usability*** to adapt to the new tendencies in the HCI community. This adaptation presents a set of facets that are included in the user experience concept such as cross-cultural, accessibility, usability, among others. (Chapter 2)

- ***The large and easy expandable repository of information*** where it is possible to save all the required data to carry out the entire process of a heuristic evaluation. The selection of the most adequate heuristics for a specific system is one of the most time-consuming tasks of the heuristic evaluation because there is not a library or repository where all the needed information is saved. Our proposal includes a large laundry of data and information that is able to provide all required information to carry out the complete heuristic evaluation process. However, in any case, the heuristic data is the most important. (Chapter 3)

- ***The semi-automatic selection of the most suitable heuristics*** for different kinds of interactive systems and for different goals depending on the type of design or evaluation required. By using the information of the repository and after running the recommendation algorithm, the list of the most appropriate set of heuristics for a specific system is given semi-automatically. (Chapter 3)

- ***Management of conflicting heuristics*** is provided to enable the solution of possible conflicts that appear when different sources of heuristics are considered in a specific interactive system. The representation of conflicting heuristics through rationale notation facilitates the comprehension of the problem and a quick and easy resolution. (Chapter 3)

- ***Quantitative and qualitative measures for presenting the UX evaluation results*** to the final client. Usually, heuristic evaluation provides qualitative data. Without ruling out qualitative information, different measures are proposed to aim for objectivity which cannot be obtained using qualitative results. (Chapter 3)

- ***Common Industry Format (CIF) reports.*** Quantitative as well as qualitative results are provided in an editable format following the *ISO/IEC 25062 Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Common Industry Format (CIF) for usability test reports*. (Chapter 4)

- ***The implementation of this methodology in a real web-based resource called Open-HEREDEUX*** (Open HEuristic REsource for Designing and Evaluation User eXperience) that permits the execution of this methodology in real projects. Therefore, considering that the definition of the methodology is the main part of the research, its implementation becoming a real resource to be used in real projects

provides the second most important contribution of this project in the HCI community. (Chapter 4)

- ***The viability study of the methodology and Open-HEREDEUX interaction* in real contexts *through an international company*** with an emergent UX department. It gives a real view into the usage of the methodology and Open-HEREDEUX. It also provides the needed added value to improve it according to the results of the study. (Chapter 4)

1.4 Structure of the document

This document is divided into 5 chapters and 3 appendices that are structured as follows:

Chapter 1 presents the introduction of the document. It details the motivation of this research, the goals and the main contributions of this thesis. Moreover, two real case studies are introduced as the base cases for all of the research.

Chapter 2 details the usability and the user experience concepts. The different definitions of each term, the existing methodologies to consider the usability and the user experience in the development process of an interactive system and the translation from the usability term to the user experience concept are described. In addition, heuristic evaluation methodology is presented and the state of the art about tools that support any step of the methodology is analyzed. Then, the different measures that are used for getting heuristic evaluation results are documented. Furthermore, different ways for documenting conflicts among guidelines are detailed and a possible notation to document these conflicts is also presented.

Finally, what is missing in each step of the methodology poses challenges for which we provide solutions.

Chapter 3 is one of the most important chapters because it proposes the new methodology that adapts and optimizes the heuristic evaluation to a new methodology based on heuristics. It also helps in the consideration of the user experience in the design and evaluation steps of the development process of interactive systems. Therefore, this chapter details the repository of information where all the needed information for considering the user experience in every step of the development process is explained. Following, the semi-automatic process of suggestion heuristics for a specific interactive system is proposed. Thereafter, we detail how to support the scoring step of the heuristic evaluation. And, in the end

(and as a result of the scoring process), different measures for quantifying the UX of an interactive system are proposed.

Chapter 4 introduces the implementation of the methodology based on heuristics in the web-based resource called: Open-HEREDEUX (Open HEuristic REsource for Designing and Evaluating the User eXperience). Then, the benefits of the resource are presented. Finally, how the resource is implemented following the user centered design methodology is detailed including the feasibility test in a real company.

Chapter 5 is the last one. It gives the conclusions and the future work of this PhD.

Appendix A presents a table where the specific authors that defined each usability category are detailed.

Appendix B lists an example of heuristics through two different visions: heuristics included in different facets and heuristics for some attributes of the ISO/IEC 25010:2011 standard.

Finally, Appendix C shows the CIF template used to report UX evaluation results.

Chapter 2

"The art of winning is learned in defeat"
"L'art de vèncer s'aprèn a les derrotes"

State of the Art

2.1 Introduction

The state of the art of this PhD includes different areas. The first two sections present the two main concepts of this project: usability and user experience. Specific definitions and methodologies for the evaluation of both concepts are detailed.

Following the state of the art about usability and user experience concepts, heuristic evaluation is introduced. The definition of the heuristic evaluation methodology is explained. Then, the state of the art about the steps that make up a heuristic evaluation is detailed. Therefore, this part of the state of the art includes: different sets of heuristic definitions, how to manage conflicting heuristics, tools for carrying out a heuristic evaluation and different possibilities for obtaining results from the heuristic evaluation.

Finally, a list of deficiencies from the heuristic evaluation methodology are listed and proposed as the challenge of this PhD.

Let's read the usability and user experience definitions.

2.2 Usability

Usability is one of the oldest terms that HCI practitioners use when they talk about the quality of interactive systems. This section presents different definitions for usability. Then, the set of methodologies to evaluate usability are described. Finally, the current results that it is possible to get from a usability evaluation are also detailed.

2.2.1 Definitions

Up to now, many definitions appear in literature for describing the usability concept. In the following, the most used definitions are presented in chronological order.

In 1993, Nielsen said that "*it is important to realize that usability is not a single, one-dimensional property of a user interface. Usability has multiple components and is traditionally associated with these five usability attributes: learnability, efficiency, memorability, errors, satisfaction"*. [NIE93]

One year later, Preece defined usability [PRE94] as "*a measure of the ease with which a system can be learned or used, its safety, effectiveness and efficiency, and the attitude of its users towards it".*

The author of the popular book called *Don't Make Me Think* [KRU05] says that *"After all, usability really just means that making sure that something works well: that a person of average (or even below average) ability and experience can use the thing - whether it's a Web site, a fighter jet, or a revolving door - for its intended purpose without getting hopelessly frustrated".*

Then some quality standards appeared. In 2001 International Standard*, ISO/IEC 9126: Software Engineering - Product Quality defined usability* [ISO9126] as "*the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions".*

In 2008, the International Standard, ISO 9241-11 2008 [ISO9241], provided guidance on usability and defines it 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*".

Finally, the most recent standard, ISO/IEC 25010:2011 [ISO25010] presents usability as the "*degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. Usability can either be specified or measured as a product quality characteristic in terms of its subcharacteristics, or specified or measured directly by measures that are a subset of quality in use".*

Therefore, in this project the definition proposed in ISO/IEC 25010:2011 is used when the term usability appears given that it is the most recent definition.

2.2.2 Usability evaluation methodologies

There are different classifications of usability evaluation methodologies depending on different features: who runs the evaluation (such as an expert analysis or with user participation), the type of the methodologies (for instance, inspection, inquiry and test methodologies), or where the evaluation is carried out (such as in the field or lab studies) [NIE94a] [DIX09].

However, in any case the headings of the classification include the same specific methodologies. Therefore, in this project the usability methodologies are presented as one of the most used and general classifications: inquiry, test and inspection usability evaluation methodologies.

2.2.2.1 Inquiry

Usability evaluators obtain information about users' likes, dislikes, needs, and understanding of the system by talking to them, observing them using the system in real work (not for the purpose of usability testing), or letting them answer questions verbally or in written form. Inquiry methods include field observation, focus groups, interviews, logging actual use, proactive field study and questionnaires.

2.2.2.2 Test

In this type of usability evaluation methodology, representative users work on typical tasks using the system (or the prototype) and the evaluators use the results to see how the user interface supports the users doing their tasks. Some examples of testing methodologies are coaching method, thinking aloud, co-discovery learning, performance measurement, question-asking protocol, remote testing, and retrospective testing, among others.

2.2.2.3 Inspection

In usability inspection methodologies, usability specialists examine usability aspects of a user interface. The most common inspection methods are cognitive walkthroughs, feature inspection, pluralistic walkthrough, perspective-based inspection, standards and the main character of this project: the heuristic evaluation.

2.2.3 Usability evaluation results

Regardless of what usability evaluation methodology is used, there are two general kinds of evaluation results: qualitative/quantitative results and formative/summative results.

Qualitative results involve analysis of data such as words (e.g., from interviews), pictures (e.g., video), or objects (e.g., an artifact) [NEI07]. Qualitative results present specific information about the quality of the interactive system that is evaluated.

Quantitative results involve the analysis of numerical data [NEI07]. Quantitative results present data about the amount of quality that an interactive system has. Table 2.1 presents a comparison between qualitative and quantitative results [NEI07].

Qualitative	Quantitative
‘‘All research ultimately has a qualitative grounding” – Donald Campbell	‘‘There’s no such thing as qualitative data. Everything is either 1 or 0” – Fred Kerlinger
The aim is a complete, detailed description	The aim is to classify features, count them, and construct statistical models in an attempt to explain what is observed
Researcher may only know roughly in advance what he/she is looking for	Researcher knows clearly in advance what he/she is looking for
Recommended during earlier phases of research projects	Recommended during latter phases of research projects
The design emerges as the study unfolds	All aspects of the study are carefully designed before data is collected
Researcher is the data-gathering instrument	Researcher uses tools, such as questionnaires or equipment to collect numerical data
Data is in the form of words, pictures or objects	Data is in the form of numbers and statistics
Subjective – individuals’ interpretation of events is important, e.g., uses participant observation, in-depth interviews etc.	Objective – seeks precise measurement & analysis of target concepts, e.g., uses surveys, questionnaires etc.
Qualitative data is more ‘rich’, time consuming, and less able to be generalized	Quantitative data is more efficient, able to test hypotheses, but may miss contextual detail
Researcher tends to become subjectively immersed in the subject matter	Researcher tends to remain objectively separated from the subject matter

Table 2.1 Qualitative and quantitative results

Formative results show very detailed results of the evaluation and inform about the specific and detailed problems and possible solutions that could be applied in the interactive system.

Finally, we present summative results to sum up the information collected in the evaluation are presented. The main goal of the summative results is to obtain a few data to show the interactive system state.

According to the goals of the evaluations, some type of results will be more suitable than others. For instance, if a list of specific improvements is needed to be able to enhance the usability of some part of an interactive system, the formative and qualitative results should be presented. On the contrary, if the main goal is the validation of an important improvement of the efficiency in terms of time per task for a specific functionality, summative and quantitative information should be provided.

The most important aspect of evaluation is to get the needed information in every situation according to the main pre-established goals.

2.3 User eXperience

User eXperience (UX) is a concept that appears as an evolution of usability. Some years ago, the quality of use of an interactive system was measured through the usability attribute. However, as technology flooded all human activities, usability and HCI practitioners realized that usability was not enough to get a quality product that provokes positive feelings in end users. Therefore, the most common term to talk about product (interactive system) quality is UX.

Due to the novelty of UX, it lacks some consensus that is presented in the following sections. First, the existing UX definitions are detailed and our UX definition is proposed. Second, the lack of methodologies for considering the UX in the development process of an interactive system is explained.

2.3.1 UX Definitions

On the one hand, literature proposes the approach where UX only highlights emotional features. For example, in Hassenzahl's definition [HAS06], only emotional aspects of the user are included. On the other hand, other authors such as Peter Morville [MOR05] believe that UX includes more aspects than emotional, thus extending the meaning of the UX concept.

Specifically, Peter Morville [MOR05], one of the distinguished authors in the UX topic, defined the "user experience honeycomb" which included: *usable, useful, desirable, valuable, credible, findable and accessible*.

The ISO 9241-210:2010 [ISO9241] standard provides one of the most distinguished definitions for UX: "*A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service*".

And the following five definitions which were collected by E. L Law et al. in [LAW09] are also used.

"Encompasses all aspects of the end-user's interaction with the company, its services, and its products. The first requirement for an exemplary user experience is to meet the exact needs of the customer, without fuss or bother. Next comes simplicity and elegance that produce products that are a joy to own, a joy to use. True user experience goes far beyond giving customers what they say they want, or providing checklist features. In order to achieve high-quality user experience in a company's offerings there must be a seamless merging of the services of multiple disciplines, including engineering, marketing, graphical and industrial design, and interface design." [NNG11]

"A consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.)" [HAS06]

"The entire set of affects that is elicited by the interaction between a user and a product including the degree to which all our senses are gratified (aesthetic experience) the meanings we attach to the product (experience of meaning) and the feelings and emotions that are elicited (emotional experience)." [DES07]

"The value derived from interaction(s) [or anticipated interaction(s)] with a product or service and the supporting cast in the context of use (e.g. time, location, and user disposition)." [SWA07]

"The quality of experience a person has when interacting with a specific design. This can range from a specific artefact such as a cup toy or website up to larger integrated experiences such as a museum or an airport." [UXN11]

Another definition was proposed by Alben [ALB96]. "*The aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it. If these*

experiences are successful and engaging, then they are valuable to users and noteworthy to the interaction design awards jury".

Finally, UX in interactive TV context was proposed by Pirker [PIR11]. *"The user experience when interacting with an iTV system in the specific living room context is mainly influenced by: the subjective perception of the quality of experience that is elicited by the interaction of a user with the interactive TV system, which may change dynamically depending on the situational context of usage and time. Factors influencing the quality of experience include feelings and emotions that are elicited (emotional experience), the degree to which our senses are gratified by the system (aesthetic experience), meanings and values that are attached to the system, the perception of system characteristics like utility, purpose and usability, and how well these factors fit the current situational and temporal context."*

Despite the fact that the definitions presented are valid in specific contexts such as in [PIR11], they do not include aspects which should be considered when evaluating UX. In some definitions such as in [LAW09] and [UXN11], the interaction context is not so clear. In [NNG11], the main topic is concerned in company aspects. Other definitions do not refer to facets such as accessibility [HAS06], cross-cultural [DES07] or adaptability [SWA07].

Facet / Author	Dependability	Usability	Playability	Accessibility	Plasticity	Emotional	Desirable	Useful	Findable	Communicability	Cross-cultural	Valuable	Credible	Context	Facets/Author
[ALB96]	x	x		x	x			x		x	x			x	8
[MOR05]		x		x			x	x	x			x	x		7
[HAS06]	x	x				x	x	x	x	x	x		x	x	10
[DES07]		x				x	x			x	x	x			6
[SWA07]		x												x	2
[ISO9241]						x									1
[NNG11]		x				x	x	x							4
[UXN11]						x									1
[PIR11]	x	x	x			x	x	x	x			x		x	9
Who considers the facet	3	7	1	2	1	6	5	5	3	3	3	3	2	4	

Table 2.2 Facets included in each UX definition

Table 2.2 displays the facets considered in each UX definition here presented. In order to fix which facets are involved in every UX definition, I have reviewed each UX definition (and all the complementary research carried out to provide each author/s with the needed information for proposing the UX definition) and checked which facets are included in it.

Bearing in mind that UX appeared to include emotional aspect in the product quality, the first main topic was to detect if UX definitions include more than emotional aspects. Up to this point, if UX definition included more than emotional aspects, other facets were detected according to the specific definition (and its related research) of each author.

2.3.2 UX evaluation methodologies

Regardless of the new tendency in UX, currently there is no specific technique to evaluate the UX when users interact with an interactive system [VER10]. Certainly, UX experts can use usability evaluation methods for evaluating the UX, because the "oldest" and the most traditional facet used when evaluating aspects related to the quality of use is usability. Nevertheless, a consensus about what is the best methodology for evaluating UX does not exist [VER10] [ROT11].

Therefore, this project is based on one of the most traditional methods to evaluate the usability of an interactive system, the heuristic evaluation methodology.

2.4 From usability to the user experience

As we mentioned in previous sections, usability was the first concept that appeared in the development process for software interfaces. However, nowadays, the trend of the UX concept makes people forget the usability term. But what is our view about what really happened?

Some years ago, there was a trend to include in usability facet several features that are more suitable in other facets. This means that usability included security, accessibility, and cross-cultural aspects, and so on. Thus, in reality what is important to clarify is the specific nomenclature. Now, the HCI community tries to name each concept with the most exact word. Therefore, the features that usability included are divided into concepts that are more specific and usability retains its original meaning.

In addition, the features related to emotions appear to be included in the quality in use such as the main aspect to obtain a positive experience when users interact with a system. However, are emotions the only facets needed so that someone enjoys a positive experience or a quality product? According to our proposal (it shall be presented later), the answer is: definitely no, there are more aspects that should be considered to provide users with as positive as possible experience.

However, in the end, it does not matter that we are talking about usability or UX. The most important thing is provide people with better interfaces to facilitate their daily life.

2.5 Heuristic Evaluation

Bearing in mind the main goal of this research detailed in Chapter 1, the *definition of a quick and economical methodology* that permits the consideration of the *user experience in different steps of the development process* of an interactive system, heuristic evaluation is the selected methodology.

The next section presents its definition and the process to carry out a heuristic evaluation. Then, the state of the art about the main aspects of this methodology is detailed to detect, in the following section, the main deficiencies of the methodology.

The state of the art is based on the three main steps of the heuristic evaluation methodology. First of all, the usability heuristic definitions are presented chronologically. Following this, the tools that help in the management of conflicting heuristics are given. Then different types of tools that support in a more or less sophisticate way the methodology are described. Finally the existing ways to score the heuristics and the type of results are detailed.

This section is based on my Master's Degree Project [MAS10], so the reader can find full details in this document.

2.5.1 Definition and methodology

Heuristic Evaluation (HE) is an inspection methodology used to evaluate the usability of interactive systems. HE was created by Johnson, Ravden and Clegg [JOH89] but it was promoted by Nielsen and Molich [NIE90] one year later.

The methodology structures the evaluation in three main steps: the organization of the evaluation, the evaluation of the heuristics and the extraction of results.

In the first step, the evaluation manager selects the list of heuristics that best fits the specific system, taking into account the goals of the evaluation. Then, the evaluation manager chooses who will be the evaluators and determines the severity factors that the evaluators will use to score each heuristic (by default and according to Nielsen's scale [NIE90], impact, frequency and persistence are the usual severity factors that are considered in the heuristic evaluation).

In the second step, each evaluator scores each heuristic individually using the severity factors selected previously. Sometimes, evaluators write down some observation to clarify their scorings. Once all evaluators have scored all the set of heuristics, the results should be reached.

The extraction of results includes two actions: the post-evaluation meeting and the extraction of results per se. The post-evaluation meeting is used to decide the final scoring of heuristics evaluated with different criteria. The meeting ends with a qualitative results report which brings together the scorings and the observations of the evaluators.

HE is one of the most used methodologies [UPA09] because of its quickness and cheapness (compared to other methods such as the user test). It can be applied in all the different steps of the development process [NIE90], it does not need exhaustive planning and the evaluation process is very intuitive. In addition, it is believed to detect 42% of serious problems and 32% of minor problems [NIE90]. For all these reasons, it is often considered as a discount usability technique [DIX09]. Nevertheless, it presents some deficiencies that induce a slower and more expensive process.

In the following section the state of the art about the existing technology to support each step of the heuristic evaluation is detailed. In the end, a discussion about the existing deficiencies of the heuristic evaluation methodology is presented.

2.5.2 Heuristics

The use of heuristics to find the solution for a problem is a very old technique. However, the first design guidelines or principles appeared in 1986. Since this year, different authors proposed many different heuristics to be considered in the

development process of an interactive system, manly website and desktop applications.

Furthermore, heuristics are widely available but in many different formats with contents varying both in terms of quality and level of details. They appear in literature under various names. For instance, Vanderdonckt [VAN99] defined five types of ergonomic sources: *design rules, guidelines sets, standards, style guides and ergonomic algorithms*. Moreover, Mariage, Vanderdonckt and Pribeanu [MAR99] extended this classification adding *principles, recommendations, isolated guidelines and user interface patterns.* Although Stephanidis and Akoumianakis [STE99] also talked about *standards and recommendations*, they introduced the term design *heuristics* as specific context-dependent guidelines applicable in specific systems.

There are fuzzy differences in some terms as style guide and design rule. [VAN99] presents a table trying to show some differences about these terms. However, the most usual classification details that *principles and standards* are more general and abstract than *recommendations, design rules or ergonomic algorithms* as [MAR99][STE99] illustrate.

> *Bearing in mind that the methodology base of this research is the heuristic evaluation method, the term* ***heuristic*** *will be used along the entire document.*

However, this does not mean that other options are ruled out. In any case where the term heuristic is used, other terms such as guideline, principle, recommendation, design rule, style guide among others, can also be considered. The research is focused on heuristic evaluation methodology but any type of guidelines can be included in it.

Figure 2.1 shows the historical schedule of different sets of usability heuristics from 1986 to 2010.

It is important to highlight that the definition of the first usability set of heuristics does not coincide with the definition of the heuristic evaluation methodology. The initial usability heuristics were recommendations that designers could use to design a specific interface, which appeared in 1986. The heuristic evaluation methodology was defined in 1990 [NIE90] and up to now, heuristics can be used as design recommendations and as aspects to evaluate in an existing interface.

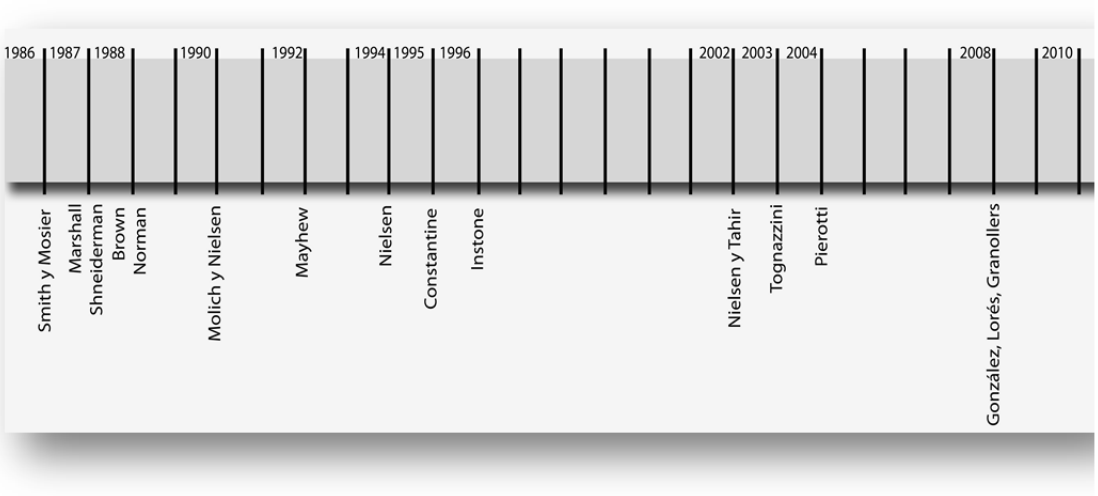


Figure 2.1 Schedule of heuristic definitions

The first usability principles were defined by Smith and Mosier in 1986 [SMI86]. They decided to write a list of guidelines that can be used for designing software interfaces. One year later, Marshall [MAR87] proposed a set of heuristics to support the human-computer interaction and Shneiderman, at the same year, presented eight golden rules for interface design [SHN87].

The following year, in 1988, Marlin Brown [BRO88] wrote the book called "Human-computer interface design guidelines" where sets of usability criteria were defined for the design of interactive interfaces. Just like Brown, Norman listed seven heuristic principles to consider in a user-centered design process [NOR88].

Molich and Nielsen suggested nine principles to improve dialog between people and computers in 1990 [MOL90]. Then in 1992, Mayhew described a set of heuristic principles related to the design of user-centered systems [MAY92]. Only two years later, Nielsen proposed categories that included a list of heuristics to cover many different aspects related to the usability of interactive systems [NIE94].

In 1995, Larry Constantine defined five general rules and six specific principles for the design of usable interfaces and to facilitate learning by end users [CON95]. One year later, in 1996, Instone presented a technical report where a set of heuristic principles for website applications was specified [INS97]. Then, in 2002, Nielsen reappears with Tahir for the presentation of the book called "Homepage usability:

50 websites deconstructed" [NIE02]. In this book, 113 heuristics are detailed to use in the design of homepages. After this definition, Tognazzini, on the website "Ask Tog", suggested a set of principles for GUI and web environments in 2003 [TOG03].

Pierotti [PIE04] extended the set of heuristics defined by Molich and Nielsen [MOL90]. He added three more categories and divided the proposal of Molich and Nielsen and his own set of principles into more specific heuristics.

Finally, González et al. presented a collection of heuristics [GON08] according to the Nielsen, Schneiderman, Instone, Tognazzini, Constantine and Mayhew proposals. The main aim of this collection was obtaining a wider usability set of heuristics to evaluate the usability on interactive systems that includes a heuristic definition of all the mentioned authors.

Apart from this specific definition, since 2008 [TRA09] [USD06] appear as web libraries where many usability heuristic definition are provided from the review of different sources. These collections are a good resource to research guidelines for specific aspects. However, it is only a public or open collection to be taken into account in the selection of the most suitable heuristics for a specific interactive system.

2.5.3 Management of conflicting heuristics

Apart from the selection of heuristics, another very important aspect in the evaluation planning is the detection and management of conflicting heuristics. This aspect occurs when different sources of heuristics are used.

> In our context conflicting heuristics are defined as the combination of different sources of heuristics that could end-up with a huge list of entries containing duplicated entries, similar statements using different terms, heuristics that refer to elements that are not relevant to the project, and potentially conflicting heuristics.

For example, security heuristics recommending validation steps those contradict with usability heuristics that recommend minimal actions. In order to design a user interface meeting both usability and security in such a context, a cleaning-up

selection process was required to provide reliable, consistent and usable set of heuristics.

The resolution of conflicts is a daunting and demanding task that often requires taking into account the trade-offs associated with alternative design choices. Therefore, whenever a good solution for solving conflicts among heuristics is found, it is worth the effort recording and documenting it for further reuse. In this section, the state of the art is presented.

These problems related to the selection of heuristics from different sources have been previously reported in literature (such as in [VAN01] and [VOG01]) and motivated the development of tools for working with guidelines [MAR02] and [VAN99]. Nonetheless, the problems related to the occurrence of potentially conflicting heuristics have been poorly documented so far.

At this point, it is important to point out that we always use the word “heuristic” but other authors use other words such as guideline, recommendation, rules, among others. In the state of the art, the preferred word of the author is used.

Several works [VOG01] [ABA01] report problems associated with the management of guidelines sets. Vanderdonckt [VAN99] discusses the potential occurrence of conflicting problems when selecting guidelines from diverse sources and he proposed a dedicated process for selecting the best set of guidelines for a specific interactive system.

Vogt [VOG01] extends that work by proposing taxonomy of 11 types of problems associated with conflicting guidelines. Abascal et al. [ABA01] explicitly mention that a step for the resolution of conflicting guidelines should be performed when selecting them for a teaching context; nonetheless they do not describe how conflict resolution can be specified.

In [CRA05] a set of unresolved problems are presented for the tools for working with guidelines. One of these specified problems is its maintenance. The authors points to the conflicting guidelines as an example of an unresolved problem [ABA01a]. Finally, the state of the art about the process for getting the most adequate set of guidelines is done and the most highlighted research in the definition of a process for getting guidelines is [VAN99]. In this research, a process for developing a set of guidelines is described using five milestones as the main point for getting a tool to work with the guidelines. Another work is [DEA06], which

defines a generic framework for the collaborative development of principles and standards involving many experts in their usage.

To sum up, despite the fact that several works agree on the existence of potential conflicting heuristics, there is not any research so far proposing a methodological approach for dealing with such conflicts. Existing tools can handle diverse heuristic sources but they are not able to show if conflicts appear between them. Moreover, even if designers are able to solve the conflicts among heuristics, they have no support to document their arguments leading to the solution, which can be lost in future projects.

As you shall see in Chapter 3, our proposal for the management of conflicting heuristics is presented using rationale notation. So, at this point, the presentation of some concepts about rationale notation is needed.

2.5.3.1 Rationale notation

The resolution of conflicting heuristics requires the systematic exploration of design options. In a previous work Lacaze et al. [LAC06] and Martinie et al. [MAR10] propose the Design Rationale TEAM (Traceability, Exploration and Analysis Method) and the tool called DREAM (Design Rationale Environment for Argumentation and Modelling) to support the systematic exploration of design options during the development process of interactive systems. Hereafter, the main concepts of the TEAM notation for describing heuristics are detailed.

TEAM notation is an extension of MacLean's QOC (Question Option Criteria) [LAC06] which allows the description of available options for a design question and the selection of an option according to a list of criteria. The TEAM notation extends QOC to record the information produced during design meetings, including:

- Questions that have been raised.
- Design options that have been investigated and the ones that have been selected.
- Criteria that have been used for evaluating the options considered.
- Requirements for the system and how they are supported by design options.
- Factors that have been taken into account and how they relate to criteria.
- Arguments and documents used to explain the design options.
- Task models corresponding to options.
- Scenarios that are used to compute the value of the criteria for each option.

Figure 2.2 shows a simple TEAM model that contains all the required elements to describe heuristics. In the example below, the requirement for the Website "provide access to data" is represented by a square. The question raised during the web site design (represented by a square with rounded-corners) indicates two possible design options (represented by circles) to grant users with access to a Website: "provide direct access" and/or "ask first for login and password".

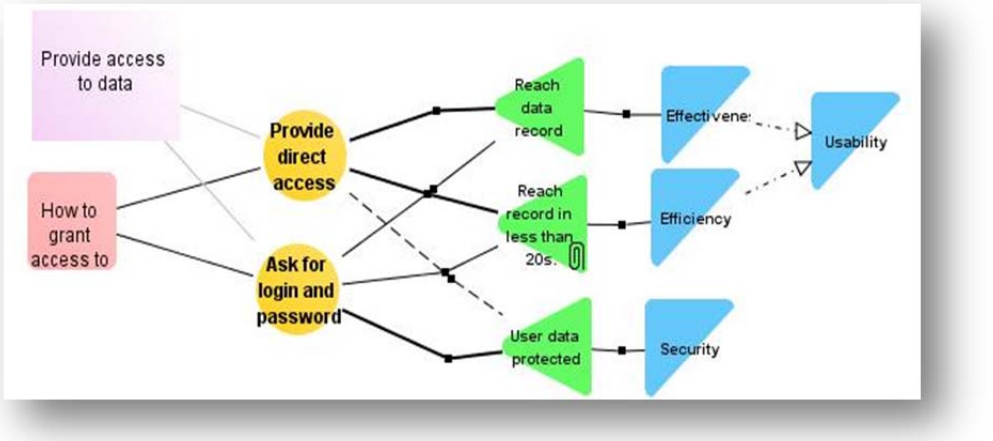


Figure 2.2 Simple model showing the main elements supported by the notation TEAM

The measurable criteria associated with design options are presented by isosceles triangles. The clip-shaped icon next to the item "reach record in less than 20s" links this criterion to the arguments and documents that can be used to measure it. The criteria can be directly connected to factors and sub-factors (represented by equilateral triangles) such as in the case of factor security and sub-factor efficiency and effectiveness that are connected to the factor usability.

The different types of lines between the criteria and options represent the fact that a given option can support (favour) a criterion (the line is bold) or not support it (the line is dotted). Thus, the option "provide direct access" supports effectiveness and efficiency but it does not support user data protection. The option "ask for login and password" strongly supports user data protection (bold line) but has an impact on effectiveness and efficiency (thin lines).

TEAM models can leverage the rationale process design by helping designers to document their decisions and choices with respect to the many options available. Moreover, TEAM models can also help to decide the reuse (or not) of design choices

when facing an already experienced issue. TEAM notation is supported by the tool DREAM which supports the edition, recording and analysis of TEAM diagrams [MAR10].

2.5.4 Tools for carrying out a heuristic evaluation

As we have presented in section 2.3.2, there are not methodologies that help in the specific process of the UX analysis. Certainly, UX experts use usability evaluation methods for evaluating UX because usability is the "oldest" and the most traditional facet related to the quality of use. Nevertheless, a consensus about what is the best methodology for evaluating (or the best combination of methodologies and techniques) UX still does not exist [VER10] [ROT11].

For all these reasons, this section will discuss tools based on the usability methodology analysis. Specifically, the state of the art is focused on tools that consider the HE as a methodology to analyze the usability of an interactive system. According to the gathered tools, the following classification is detected:

- Tools for working with guidelines.
- Questionnaire management tools.
- Tools for carrying out heuristic evaluations.
- Other tools

2.5.4.1 Tools for working with guidelines

Tools which are able to ease the management of different heuristic classifications exist. The main functionalities of this type of tool are the creation of one or more heuristic classification to manage these heuristics in an easy way and according to the user profile that will use them.

- Vanderdonckt created SIERRA (System Interactive for ERgonomic Realization of Applications) [VAN95] to represent the organization model of guidelines that he designed. The main aim of this tool was to show the entity-relation model created using a graphic interface. This graphic interface shows the Smith and Mosier [SMI86] guidelines through two different points of view: on the one hand, it shows design guidelines and, on the other hand, it presents guidelines to be used for evaluation purposes. SIERRA is only a repository where designers and evaluators can find guidelines for their specific purpose: design or evaluation.

- Sherlock is another tool designed for working with guidelines. It was developed by Grammenos, Akoumianakis and Stephanidis [GRA00] and its main

functionality is to provide guidelines for the design phase of an interactive system. This means that these tools are useful for designers and/or analysts, and also for usability experts. At the end of the interface validation, it provides graphical interpretation for the results of each evaluation. Therefore, it is also a library where it is possible to search for guidelines but only for the design step of the development process.

- GUIDE (Guidelines for Usability through Interface Development Experiences) was created in 2000 by Henninger [HEN00]. Its main goal is the creation of usability guidelines and maintenance. GUIDE provides a set of relations between the guidelines to make up a guideline hierarchy according to Smith and Mosier guidelines [SMI86]. The tool also allows extending, updating and deleting the guidelines and the relations among them. In the same way as the two previous tools, GUIDE is nothing more than a repository that stores guidelines that users can select manually for a specific case.

- A tool was developed as part of a research project in the Master's Degree called "Human-Centred Computer Systems (MSc)" from the University of Sussex [SUS09]. Its main aim is to help with the challenging task of choosing the needed guidelines to be used in carrying out heuristic evaluations for different devices (websites, mobile phones, pocket PCs, tablet PCs, interactive kiosks, interactive TVs, personal organizers and interactive toys). It uses Nielsen's heuristics and some adaptation to cover the features of these interactive systems. The tool is free and can be used for usability experts and for pedagogical reasons. However, it is not possible to configure the exact aspects of the interactive system that you would like to evaluate. The evaluation manager can only select the system and receive the list of guidelines but without any refinement or previous selection of criteria to get more suitable results. In addition, the types of interactive system are limited to the above-mentioned ones.

2.5.4.2 Questionnaire management tools

Another kind of tool consists of those that are presented as questionnaires. Veeut and Gedis guides are two examples of this type of tool.

- Veeut (Virtual Environment Evaluation Usability Tools) [COL09] is a collaborative tool that enables carrying out usability evaluations such as cognitive walkthroughs, user tests or heuristic evaluations. If the heuristic evaluation technique is considered, Veeut allows for a manual selection of a set of heuristics (bearing in mind that heuristics are a set of questions that should be answered) and the scoring of the preselected heuristics (the rank of the scores is also possible to preselect according to a numerical rank, open, close or mixed answers). However, Veeut does not present any kind of results.

- The main aim of the Gedis guide is the realization of security audits in industrial interfaces [PON06]. The tool stores the set of questions that the Gedis guide presents. However, new questions can be added. The administrator of the evaluation can manually choose the best guides or questions for the specific industrial interface but the list of guides is always the same. It also defines the severity factors. Finally, only one evaluator can answer all the questions providing the final evaluation.

2.5.4.3 Tools for carrying out a complete heuristic evaluation

This section presents tools that try to support some parts of the heuristic evaluation process such as the manual selection of guidelines or the scoring process. Specifically, details about R-IDE, UsabAIPO-GestorHeuristica, Accusa and Gestor-Heuristicas.

- R-IDE is available on the Internet [KEM06]. The process starts when the evaluation manager selects specific attributes to define the most adequate set of heuristics. These attributes are: the type of system to be evaluated, the general category, the specific category and the user group. According to these attributes, the tool provides the most adequate set of heuristics proposed by Nielsen and a customized set for e-commerce (which can be manually refined). It is also possible to define the severity factors that after this configuration, evaluators will use in the scoring of heuristics. To sum up, R-IDE allows the manual selection of predefined heuristics for only one interface. It does not provide evaluation managers with any kind of report for the results of the evaluation.

- UsabAIPO-GestorHeuristica manages the process of the heuristic evaluation [GON06]. According to a set of predefined heuristics (by González [GON08]), this tool provides these heuristics through a usable interface. Nevertheless, the selection of the heuristics, the severity factors and the extraction of results are possible if the user has knowledge on MS Access.

- Accusa is another tool for carrying out heuristic evaluations [SQU09]. It defines heuristic evaluation and evaluators who will score these heuristics. Both tasks are presented to be carried out manually. It is also possible to carry out the process of scoring heuristics and it provides a report with the qualitative usability problems as the evaluation results.

- "Gestor heuristicas" is divided in two modules [MUR07]. The first one is responsible for managing projects, evaluators and the set of predefined heuristics to be used. Delivering final results is also a task of this module. The second, called the evaluator module, is used by the evaluators to score the heuristics and for sending the information to the responsible module. Finally,

the responsible module extracts the list of violated heuristics and/or the report on all the scores done by the evaluators. This tool uses González's set of heuristics.

2.5.4.4 Other related tools

The last type of tool described in this section includes some other tools that are more or less related to the usability analysis of interactive system. Although the tools here presented are not really focused on the heuristic evaluation methodology, they present relevant aspects for this research.

- SUIT is a tool that is a little bit different. 2006.
- [ARD06]. It was created to carry out usability evaluation but using an extended version of the heuristic evaluation. The evaluator should review the interface following a set of predefined design patterns to detect usability mistakes. In addition, the evaluator scores each problem detected in the interface using Nielsen's severity factors [NIE90]. The most important feature of this tool is that it provides a forum where the evaluators can discuss the encountered problems (and differently scored).

- Another different resource in a MS Excel format for running a heuristic evaluation is presented by Olga Carreras in her blog [CAR11] based on the Sirius project. In her proposal, a review of the most important usability definitions [NIE90][SHN87][CON95][INS97][TOG03][PIE04][GON08] is presented. Then the set of heuristics is marked through the compliance of the heuristics and a scale from 0 to 10. Finally, the percentage of usability is presented according to the type of website analyzed.

- Finally, Userplus is a company ([SPI12]created by Lonneke Spinhof, but currently shut down as he informs by means of his Twitter account) that presents different tools to check the usability of an interactive system, basically websites or software products where it is possible to get a screenshot. It presents two different tools: *The tester tool*, used to carry out remote user tests, and the *advisor tool*, that helps the evaluator to get a usability analysis. The evaluator can upload some screenshots of the interface. Then the *advisor* automatically chooses some guidelines according to a set of predefined patterns that are detected in the screenshot. The evaluator can score each guideline (using the answers: Yes, No and N/A). In addition, the guidelines are divided using a degree of importance (very high, high, medium and low). Finally, it provides quantitative results according to the percentage of affirmative answers.

2.5.4.5 Comparison of tools

In the previous subsections, many tools have been presented. These tools help to carry out complete or some parts of the heuristic evaluation. In this section, a comparison of these tools is presented to detect the main common deficiencies.

Regarding the permitted actions on the interface, although some tools enable the creation, edition and elimination of heuristics, these actions are very restricted. Furthermore, only a few tools present a wide set of heuristics and the possibility of extending the set presented and relating to some interactive systems. Table 2.3 shows this comparison.

Tools	Create	Update	Delete	Hierarchy	Authors	Heuristics for	
						Designers	Evaluators
Sierra	N/A[1]	N/A	N/A	Yes	Smith and Mosier	Yes	Yes
Sherlock	N/A	N/A	N/A	Yes	N/A	Yes	No
Guide	Yes	Yes	Yes	Yes	Smith and Mosier	Yes	Yes
Interactive Heuristic Evaluation Toolkit	No	No	No	No	Nielsen and adaptations	No	Yes
Veeut	Yes	Yes	Yes	No	N/A	No	Yes
GEDIS guide	Yes	Yes	Yes	Yes	GEDIS guide	No	Yes
R-IDE	No	No	No	N/A	Nielsen and adaptations	No	Yes
UsabAIPO-GestorHeuristica	Yes	Yes	Yes	N/A	González	No	Yes
ACCUSA	No	No	No	N/A	N/A	No	Yes
Gestor heuristics	Yes	Yes	No	N/A	González	No	Yes
Advisor-UserPlus	Yes	Yes	Yes	N/A	N/A	No	Yes
SIRIUS	No	No	No	Yes	Review	No	Yes
SUIT	No	No	No	No	N/A	No	Yes

[1] N/A: Not available

Table 2.3 Comparison of tools

The internal architecture to store the heuristics is sometimes like hierarchy architecture and sometimes heuristics do not follow any type of structure. Moreover, even though Nielsen's heuristics are the most used in usability evaluations, few tools consider these heuristics. The tools most often use the heuristics from Smith and Mosier [SMI86] or from González et al. [GON08] since these sets are more useful than Nielsen's heuristics.

In general, heuristics are for evaluators. It means that the usability is more considered in the evaluation stage than in the design phase of the development process. If heuristics are proposed in different stages of the development process [FAR01] [VAN99a], the interactive systems will be more useful from the beginning of the development process.

2.5.5 Heuristic evaluation results

The type of results reported in heuristic evaluations depends on the severity factors used in the evaluation, the goal of the evaluation and the receiver of the results. Traditionally, the heuristic evaluation methodology uses the severity factors defined by Nielsen [NIE90]:

- The impact of the problem if it occurs: Will it be easy or difficult for the users to overcome?
- The frequency with which the problem occurs: Is it common or rare?
- And the persistence of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?

In addition, Nielsen also defined the rating scale to score each one of the severity factors:

- 0 = I don't agree that this is a usability problem at all
- 1 = Cosmetic problem only: need not be fixed unless extra time is available on project
- 2 = Minor usability problem: fixing this should be given low priority
- 3 = Major usability problem: important to fix, so should be given high priority
- 4 = Usability catastrophe: imperative to fix this before product can be released

Therefore, according to the severity factors and the rating scale, qualitative results are given through the list of problems that should be improved in the evaluated interactive system.

In literature, most projects that use heuristic evaluations as the methodology to reach usability problems, present qualitative results as the list of violated heuristics in the interactive system and some observations about it [COL09] [PON06] [SQU09] [MUR07] . 2006.

[ARD06]. However, some efforts are made to acquire quantitative results such as the amount of heuristics that are a problem [CHE05], the percentage of heuristics that

are not a problem [ALB96][ALL06], using some other statistic software [KIL09] such as ExpertChoice [CHOI86] or Saaty Scale [SAA80] or through a different formula as in [GON06] [GON09a].

Another related factor when talking about results is how the heuristics of the evaluation are scored. Traditionally, professionals and researchers follow the severity factors defined by Nielsen [NIE90] but very often, the persistence is omitted for unknown reasons.

Six real experiences are detailed as the state of the art on result extraction. In view of the fact that qualitative results are widely used, the state of the art on heuristic evaluation results is more focused on experiences that also present quantitative data. Specifically, the process to carry out the heuristic evaluation, the severity factors and the rating scale used and, finally, the final type of results are presented.

- Chen et al. carried out a usability analysis through the heuristic evaluation methodology of four electronic shopping sites [CHE05]. They used Nielsen's heuristics and a specific other set of heuristics that they needed to define since Nielsen's heuristics did not cover all the usability features of this kind of application. The severity factors and the rating scale used were those defined by Nielsen. The process to carry out the heuristic evaluation was different. Five evaluators detected the compliance of heuristics through a free-flow inspection and a task-based inspection. Once evaluators detected the violated heuristics, they scored each one using the severity factors and the rating scale previously cited. In reference to the type of results, qualitative and quantitative results were extracted. The qualitative results were observations and comments raised by the evaluators during the scoring process. The quantitative results were counted up according to the amount of violated heuristics selected in the free-flow and task-based inspection part.

- One year later, in 2006, Allen et al. presented a heuristic evaluation of a paper prototype from a medical website application [ALL06]. For this research, the considered heuristics were those defined by Nielsen [NIE90] and Shneiderman [SHN87]. Therefore, the final set of heuristics included fourteen heuristics. The usual process was used to run the evaluation. However, the used severity factor was the compliance or not of the heuristics through the rating scale proposed by Nielsen. The results were given as a list of one hundred usability problems and as a percentage of violated and non-violated heuristics. In addition, some graphs about these measures were presented to clarify the understanding of the results.

- Another real heuristic evaluation project was UsabAIPO [GON06]. It used the González's et al. heuristics and the evaluation only considered the violation of the set of heuristics using 0, 2 or 4 as the values of the rating scale. This project provided qualitative results as the set of violated heuristics and quantitative results through a statistical form to get the percentage of problems that the interactive system has.

- Bell laboratories developed in 2007 its heuristic evaluation methodology [COY07]. They made use of Nielsen's heuristics but redefining the set in eight new categories. The scoring of the heuristics was absolutely qualitative: high severity, medium severity and low severity. They used a spreadsheet to carry out the evaluation. After the post-evaluation meeting, the evaluation manager wrote the final report including only the needed improvements to be applied in the interactive system.

- Two years later, as the result of my final degree project, a usability analysis of websites from Catalan City Councils with a population less than 1000 was carried out [GON09a]. The heuristics selected where those created by González et al. The severity factors were impact and frequency, without considering the persistence. In addition, the rating scale was extended from 0 (this is not a problem) to 5 (this is not applicable in this case). Qualitative results were obtained by means of the observations written by the evaluators. Quantitative results were reached using a mathematical formula.

- The last real experience detailed in this document was developed by Kiliç and Gungor in 2009 [KIL09]. The heuristic evaluation is carried out in a usual way but with some modifications. The evaluation used Nielsen and Molich heuristics and five evaluators performed the evaluation. The first step of this evaluation was the detection of the compliance of the heuristics that is carried out by the evaluators. Once the violated heuristics were detected, each evaluator filled in the values of the impact and frequency using Saaty's scale [SAA80]. Then, the information was inserted in the ExpertChoice software [CHOI86]. This software permits the extraction of qualitative and quantitative results using a decision-making process.

2.5.5.1 Discussion of the extraction of results

The summary about the state of the art of the extraction of results is presented in Table 2.4. The main differences are in the set of heuristics selected in each project, the severity factors used, rating scale used and the type of results obtained. In four of six cases the heuristics are those defined by Nielsen (even though the set is widened or adapted to cover all usability aspects) because the interactive systems are web-based or desktop applications.

Interactive system	Heuristics	Severity factors	Ranks	Results		Statistical forms
				Qualitative	Quantitative	
[CHE05]	Nielsen and Muller	Impact, frequency and persistence	0-4	Yes	Yes	Amount of heuristics that are a problem
[ALL06]	Nielsen and Shneiderman	Compliance	0-4	Yes	Yes	% of compliance
[GON06]	González et al.	Compliance	0, 2, 4	Yes	Yes	Statistical form
[COY07]	Nielsen adaptation	Compliance	High, medium, low	Yes	No	-
[GON09]	González et al.	Impact and frequency	0-5	Yes	Yes	Statistical form
[KIL09]	Molich and Nielsen	Impact, frequency and persistence	Saaty Scale	Yes	Yes	ExpertChoice

Table 2.4 Comparison of the extraction of results

Regarding severity factors, usually researchers and professionals only make a score if the heuristics are applied or not in the interactive system ruling out the severity factors defined by Nielsen. At other times, the set of severity factors is simplified only with two of them: the impact and the frequency. In addition, the rating scale to score each severity factor is used depending on the goal of the evaluation or the kind of results that it is expected to reach.

The process of qualitative results extraction presents a significant feature. In the case that quantitative results are presented, qualitative results are also provided. However, this does not happen to the contrary, if qualitative results are obtained, quantitative ones are not always detailed.

When quantitative results are presented, the methodology to get the usability level is not always the same. There are statistical formulas, percentages of the amount of compliance and the usage of other types of applications that helps through a decision-making process. That highlights the efforts needed for standardizing the quantitative results of heuristic evaluation.

In addition, and as seen in literature, usually qualitative results are presented as formative results through a specific list of the problems that the interactive system has. Quantitative results are shown as summative results because quantitative ones gather the main information with little data.

2.5.6 Deficiencies of heuristic evaluation

Although heuristic evaluation is considered as a discount evaluation methodology, it has some insufficiencies that induce a slower and more expensive process. Bearing in mind the three main steps of the heuristic evaluation methodology, the deficiencies are in the selection of the more suitable heuristics for a specific interactive system, the manual scoring process and the results extraction. The next section details what is lacking.

2.5.6.1 Selection of the heuristics

Nowadays, the selection of the most appropriate heuristics for a specific interactive system is a manual process in two senses. The former is the existence of many little sources of heuristics. Usually, the individual who would like to carry out a heuristic evaluation needs to review the literature or use some of the tools here detailed, select different sources of heuristics and choose manually the most appropriate sets to determine the best heuristics for a specific interactive system. Therefore, the first deficiency is the non-existence of a repository or library where many (or most) heuristics are saved.

The latter is the criteria used to decide the most suitable heuristics. Once different sources of heuristics are selected, it is essential to refine the selection choosing the more specific and appropriate ones for the specific case. Up to now, this process is manual and it is based on the professional experience of the person who is preparing the evaluation. Therefore, there is not a methodology that allows the selection of the best heuristics for a specific interactive system. Furthermore, whilst this selection has been operating, some conflicts among heuristics could appear if different sources of heuristics are used. Therefore, at this point it is very important to get a list of heuristics as appropriate as possible and without any type of contradiction among the set of selected heuristics.

These two deficiencies are the most difficult and time-consuming tasks of the whole evaluation process [LAW04], because there is not a repository that contains a large set of heuristics and there is not an automatic process that selects the most adequate heuristics for every specific case.

2.5.6.2 Evaluation support

Once the most suitable sets of heuristics are selected for a specific interactive system it is time for the evaluation. Up to now, there are two different possibilities that support the evaluation process (even though, in any case, supports are really scarce).

The first possibility for carrying out the scorings is the traditional and most basic way: the paper-based option. The heuristics are printed on the paper and evaluators write the scores by hand. It is obvious that this kind of process is very rudimentary. In addition, after the scoring process, the extraction of results is also very difficult because the union of the different evaluations is entirely a manual process.

The second way to get results is using a spreadsheet such as Microsoft Office Excel or Access, Open Office Calc or Lotus Notes. Its advantage is that the information can be managed, enabling more effective results. The disadvantage is that evaluators cannot edit the same document or if they should do it, it can run at the same time because these tools do not permit collaborative work.

In brief, there is not any system that supports the complete process of scoring heuristics that facilitates the previous step, the selection of heuristics, and the next one, the extraction of results.

2.5.6.3 Results extraction

The last step of the heuristic evaluation methodology is the extraction of results. Two main deficiencies or needs appear at this point. The first is the post-evaluation meeting support and the second the automation of the extraction of qualitative/quantitative and summative/formative results.

If there were a scale concerning the difficulty of the steps of the heuristic evaluation methodology, the most difficult and time-consuming task would be the selection of the best heuristics for a specific interactive system. The second position would be for the post-meeting evaluation.

As we detailed in the explanation of the heuristic evaluation methodology, after the scoring of heuristics, the evaluators should discuss the scores that they marked different for the same heuristic. This process is so hard and time-consuming because in any case, the selection of one value of the rating scale or another one is a process that is a little subjective. Therefore, it is very difficult that all evaluators choose exactly the same values for a specific heuristic. For this reason, the list of heuristics for reviewing in the post-evaluation meeting is large. Literature does not show any solution that facilitates or supports the process.

Finally, the last deficiency appears in the process for obtaining the evaluation results. HE traditionally presents qualitative and formative results according to the evaluator's scorings and their observations during the scoring process. Nevertheless, sometimes, quantitative and summative results are achieved because

the scientific community aims for objectivity [NEI07] which cannot be obtained using qualitative results. Furthermore, if some kind of numerical results can be achieved, the standardization of the UX evaluation would be possible.

Apart from the type of needed results, nowadays the process to get qualitative results is also completely manual. Moreover, in the few cases that quantitative results are presented, these are calculated using any kind of statistical calculator. However, there is not a specific process that permits the extraction of results in a more or less automatic and standard process.

2.6 Conclusions

Bearing in mind the state of the art here presented and the deficiencies that the heuristic evaluation methodology has, the need to improve this methodology and get a better cost-effective technique appears.

In summary, the main deficiencies of the heuristic evaluation methodology are:

- The location and selection of the most appropriate heuristics for a specific interactive system is a manual process.
- There are two ways to carry out the heuristic evaluation per se: paper-based option and using a spreadsheet such as Microsoft Office Excel or Access, Open Office Calc or Lotus Notes.
- The cumbersome and manual extraction of results from the evaluation.
- Usability is the main considered facet of the heuristic evaluation methodology.

Therefore, the next research offers different challenges for each step of the heuristic evaluation methodology. These challenges are the key aspects of this PhD project:

The **location and selection of the most suitable heuristics is solved through a repository** where the needed information is stored **and through a heuristic suggestion process** to get the set of **the most appropriate heuristics in a semi-automatic process**.

Then, the process to **score the heuristics** for a specific interactive system is studied to provide the **needed management to facilitate the use** of the information saved in the repository and **the automatic extraction of results**.

Finally, **the post-evaluation meeting is omitted using statistical resources.** In addition, **automatic qualitative, quantitative, summative and formative results are given through a standard document** to report the results of evaluation.

Moreover, the **methodology will be adapted towards the user experience** concept. The next chapter details this adaptation.

Despite the fact that some deficiencies of the heuristic evaluation methodology are detailed in this chapter, there is a disadvantage of the heuristic evaluation that is not considered: the false positive problems. As is well-known and documented [JEF91], sometimes heuristic evaluation results present some problems that are not detected using another evaluation methodology (for instance user test). However, not detecting these problems using another method does not mean that the false problems do not exist [SAU12] [SAU12a]. This disadvantage is not considered in this project for two main reasons:

- Our proposal's main goal is to achieve a quick and cheap methodology to include UX in the design and evaluation steps of the development process of an interactive system. Therefore, just like other methodologies, this one also has some disadvantages, but this project's priority is to find a discount methodology.

- In addition, this disadvantage could be solved including another methodology in the development process. Our methodology is not presented to be used alone. As we have mentioned in the previous point, it is presented as a discount methodology that could be applied in different steps of the development process but another kind of methodology (preferably user test) should be applied in later steps to validate the whole development process and detect more specific problems in the real context of use.

Chapter 3

"Desperate times call for desperate measures"
"A grans mals, grans remeis"

UX methodology for the design and evaluation of interactive systems

3.1 Introduction

Although heuristic evaluation is heavily used by HCI practitioners, technological resources that support each step of the methodology are scarce. As of today, heuristic evaluation is a helpful technique that is carried out manually. It helps UX professionals in the development process of interactive systems.

This chapter presents the main research of the PhD: How to optimize the heuristic evaluation methodology to become a more cost-effective evaluation technique and how to adapt it to provide a methodology that includes UX in the development process of an interactive system.

Considering the main steps of the heuristic evaluation methodology and the state of the art presented in Chapter 2, the optimization and adaptation of the heuristic evaluation is presented through four main challenges:

- **Repository of information**: the needed data for the optimization of the heuristic evaluation methodology is detailed. Heuristics, UX facets, attributes of the ISO/IEC 25010, functionalities, components, features and the semantic relations among them are described. Apart from the information ontology, the most important parts of repository are the presentation of a new UX definition and the UX specification through attributes of the ISO/IEC 25010:2011 [ISO25010].

- **Heuristic suggestion**: the process to obtain the most suitable set of heuristics for a specific interactive system is presented. This process includes options such as different goals of the evaluation or design, different types of heuristics

regarding the receiver of these heuristics, the chance of applying financial constraints and how to document conflicting heuristics.

- **Execution of the evaluation**: This part supports the process of scoring heuristics; it supports the heuristic evaluation per se.

- **Results of the evaluation**: the results of the evaluation are presented through different UX measures as the correlation among evaluators, the automatic classification of problems and the percentage of UX degree. All these measures are proposed automatically and are reported in a standard and editable format.

3.2 Repository of information

According to the state of the art, we point out that the most valuable data is the collection of needed heuristics. However, besides the heuristics, more data is needed to be able to obtain the most suitable set of heuristics for a specific system (to be developed or evaluated). Therefore, UX facets, ISO/IEC 25010:2011 attributes, components, features, functionalities and interactive systems are defined in this section as well as the main information needed to enhance the first step of the heuristic evaluation methodology: the semi-automatic selection of the most appropriate heuristics for a specific interactive system.

3.2.1 Heuristics

The first set of usability heuristics considered in our project was collected in an exhaustive review where all usability definitions from 1986 to 2010 were reviewed.

The sets of usability heuristics considered in the exhaustive review were proposed by Smith and Mosier [SMI86], Marshall [MAR87], Shneiderman [SHN87], Brown [BRO88], Norman [NOR88], Molich and Nielsen [MOL90], Mayhew [MAY92], Nielsen [[NIE94], Constantine [CON95], Instone [INS97], Nielsen and Tahir [NIE02], Tognazzini [TOG03], Pierotti [PIE04] and González et al. [GON08].

The main goal of the review was to categorize usability. For this, we were looking how previous authors did it. Then we decided to include in our work those categories used by at least three authors. Therefore, the initial set of usability categories includes 16 ones:

Consistency, feedback, errors management, less memory load, flexibility, dialogs, user control, ease of use, short cuts, help, navigation, protection, emergency exits, search, internationalization and content.

The table showing more details about the specific categories and the authors who defined them can be found in **Appendix A**.

Another relevant result of the analysis was to determine the domain where these definitions were applicable. The heuristic definitions considered in the review were defined to evaluate a specific interactive system. Mainly all heuristic definitions can be used to evaluate website systems and/or desktop applications. For this reason, we think that we should extend these 16 categories for adapting heuristics to a specific or different kind of interactive system, since they are not useful in all contexts.

At that moment, only usability categories were detected. If our main goal is to carry out a heuristic evaluation, heuristics should be determined. Then, the next step was to select, for each category, those heuristics considered by more than one author. So, our first initial set of heuristics was comprised of 16 categories that included more than 250 heuristics.

Hereafter, a validation of this set of usability heuristics was carried out. The following section provides more details about this validation.

3.2.1.1 Heuristic proposal validation

The validation of this new heuristic proposal is presented with a comparison of heuristic evaluation results using two different sets of heuristics. The first set of heuristics includes the 16 categories and their specific heuristics presented in section 2.5.2. The other set of heuristics was defined by González et al. [GON08] since we have used it until now in different projects.

The experiment consists of two evaluations of three website systems. The first time, the websites were evaluated with one set of heuristics provided by our methodology; and the second time, heuristic evaluations was carried out using the other group of usability principles.

The evaluation sets out two main goals. The first goal concerns a comparison between both sets of heuristics to detect which set finds more usability problems.

The second goal is the validation of the new set of heuristics to detect possible improvements in their application and the needed improvement in understanding.

The websites evaluated in the validation are:

- IRB: Biomedical Research Institute of Lleida (www.irblleida.org)
- ITL: Technologic Institute of Lleida (www.itl.cat)
- Innopan: Center of Baking Technology(www.innopan.com)

Regarding heuristic improvement and understanding, our research question is that the type of sentence used in heuristics (declarative or interrogative) is very important for understanding the meaning of each heuristic and making the evaluators' work easier. At the same time, these sentences should be related with severity factors for the same reason commented above. Therefore, the proposal presents the heuristics for evaluators as questions and the heuristics for designers as declarative sentences.

In the following two subsections more details about both sets of heuristics and the process followed to achieve both goals are presented.

3.2.1.1.1 Sets of heuristics

To validate our new set of heuristics we needed to compare it with another one. On the one hand, we used a well-known heuristic that we had already used in previous works [GON08] and, the other set was our new set of heuristics detailed above.

The well-known set of heuristics is composed of 14 categories based primarily on Nielsen's ten rules but adding four more categories. We thought these four categories were necessary because some of the contents of the websites were not evaluated when they were an important part of these. Fourteen categories present 82 heuristics, in total.

According to our second goal concerning heuristic understanding, another feature about this well-known set of heuristics is that these heuristics are written in declarative sentences.

Our new set of heuristics is composed of 16 categories. Altogether, 250 heuristics comprise our new group of heuristics. These heuristics are written in interrogative sentences. We want to highlight that our proposal has two more categories than well-known sets. These categories are *protection of information and emergency exits*.

In addition, there was a little group of heuristics in both sets that we have not used because they are not suitable for the websites to be analysed. Due to our systems being simple websites, I discarded heuristics for desktop applications.

3.2.1.1.2 Carrying out the evaluations

The standard heuristic evaluation procedure was run to carry out the evaluation:

- *Planning the evaluation*: Firstly, the best heuristics for each website and severity factors were defined. Severity factors were impact (serious problem, minor problem or it is not a problem) and frequency (always, sometimes or never). This step was done for each set of heuristics.

- *Heuristic Scores*: Three evaluators carried out a heuristic evaluation with each one of both sets of heuristics. Therefore, each evaluator did 6 evaluations. In one evaluation, evaluators had to score the set of well-known heuristics. In the other evaluation, evaluators had to score our new set of heuristics divided into 16 categories. Furthermore, evaluators could write down some observations in every heuristic to detail more aspects about specific features of the website.

- *Discussion*: Finally, a meeting to discuss different scores in both evaluations to obtain a final consensus was arranged. Then, qualitative and quantitative results were extracted.

- Taking advantage of the opportunity, I asked for the heuristic's quality. Therefore, evaluators completed a satisfaction survey to provide me with qualitative information about the quality of both sets of heuristics. The next section expands the results of the experiment.

3.2.1.1.3 Quantitative and qualitative results

After carrying out the heuristic evaluations detailed above, results can be obtained. However, results about the evaluated websites are not presented; the sights are focused on the heuristic definition quality.

Quantitative results are specified through a comparison of a number of problems that evaluators found in websites using the above-mentioned sets of heuristics. Then, qualitative results were extracted from satisfaction surveys that evaluators filled in after performing all heuristic evaluations using both sets of heuristics.

Table 3.1 shows the number of "minor problems" that evaluators found using our sets of heuristics (Remember: we say "Well-known" for "traditional" heuristics and

"New heuristics" for the new set of usability heuristics acquired in the exhaustive review from 1986 to 2010).

Web Page	Well-known heuristics	New heuristics	Difference
IRB	9	15	6
ITL	8	18	10
Innopan	10	15	5

Table 3.1 Minor problems

The second table, Table 3.2 presents the number of serious problems obtained.

Web Page	Well-known heuristics	New heuristics	Difference
IRB	13	17	4
ITL	13	30	17
Innopan	13	36	23

Table 3.2 Serious problems

Finally, Table 3.3 details the number of total problems, minor and serious, that we found using both sets of heuristics.

Web Page	Well-known heuristics	New heuristics	Difference
IRB	22	32	10
ITL	21	48	27
Innopan	23	51	28

Table 3.3 Total of problems

The second main goal of the validation is to validate the heuristic quality through the detection of possible improvements in their application and its understanding. Thus, the evaluators' impressions were gathered, as users of heuristics, to improve heuristics. Therefore, a satisfaction form was prepared to evaluate this aspect with these main topics:

- Heuristics coverage.
- Heuristics understanding.
- Heuristics suitableness with answers of severity factors.

Concerning the first topic, evaluators pointed out that the new set of heuristics is more complete than well-known sets of heuristics. Additionally, two evaluators stated that the new set can detect more usability problems than well-known sets of heuristics. Even so, one evaluator thinks that both sets have the same number of usability deficiencies.

However, when we asked them if they were sure that these sets of heuristics covered all the usability problems in these websites, the answers were different. One evaluator assures that the heuristics cover all usability problems because the evaluated sites are very simple. Another one says that with a union concerning both sets it is possible because the sets are complementary. The third evaluator is not sure that all heuristics cover all usability problems.

In reference to our second goal about heuristics understanding, evaluators told us that they prefer to use interrogative sentences because they are more intuitive for answering a question than declarative sentences.

They think that we must rewrite some heuristics because sometimes they are a little confusing. Finally, the evaluators disagreed with severity factors. Everybody stated that the options of severity factors were not suitable for answering heuristics in a comfortable manner.

3.2.1.1.4 Discussion

According to the quantitative results, we trust in our new set of heuristics due to it detects more problems. However, these results are obvious if we consider that the number of heuristics in our new group of heuristics triples the number in the well-known set of heuristics.

When Nielsen and Molich defined their usability principles [NIE90], they tried to compress other large usability definitions because they wanted to present a usable set of usability principles that people could use to easily evaluate the usability of interactive systems. But now, we need to extend these usability principles because we consider that their usability principles do not cover all new features or components of interactive systems that we were working on. Therefore, we have only added new heuristics because we wanted to obtain a minimum set of usability heuristics to find all usability problems in any interactive system.

In reference to the number of problems, with regards to minor problems, our new set of heuristics detect more problems than well-known groups of heuristics. The same appears with serious problems; our new set of heuristics discovered more usability mistakes than well-known sets of heuristics. It is obvious that problems doubled in two cases, as can be seen in the tables above.

Furthermore, our heuristics can be enhanced. Evaluators liked our new set of heuristics more (than the well-known group) because it finds more usability

problems and it is more intuitive for scoring a question than scoring a declarative sentence. However, they suggested some improvements, such as every heuristic must include an additional description for improving heuristic understanding. Moreover, it is necessary to change severity scores because it is very difficult to score every heuristic with the rating scale used in this project.

As a conclusion, the new group of heuristics is better than the well-known set but it is necessary to rewrite some heuristics to enhance their understanding.

3.2.1.2 New heuristics from the case studies

Up to now, I have dealt with a set of validated heuristics for website and desktop applications. Now, it is time for including the heuristics used in the case studies presented in section 1.1. In these case studies, two different interactive systems were evaluated: virtual assistant and public kiosk. In both cases, we could not use the same heuristics that we usually use for evaluating websites because the main goals of the evaluations were very different from the usual ones. Thus, I had to collect manually the most adequate set of heuristics for these systems and it was more complicated than I had thought.

In the first case study, I reviewed literature to collect heuristics for virtual assistants and I did not find a specific set of heuristics which covers facial expression and dialog problems. However, four sources were found that I could mix up to gather a complete set of heuristics for covering all features of virtual assistants. These studies concern question-answer systems [MAR07], dialog management in virtual assistants [ROD02], best practices for speaking avatar design [SIT03] and emotional heuristics [LER08].

Although these studies do not present heuristics themselves, I could extract information to consider it for creating heuristics for our particular features in the virtual assistant. Table 3.4 shows the new set of heuristics for virtual assistants.

In Citizen's Information Points (CIPs), many works were considered to collect a complete set of usability and accessibility heuristics for covering all features of this kind of interactive system. Literature about design and evaluation in public kiosks [MAG97] [NIE02a] [GUT07] [EVO04] and design and evaluation in general [GRA05] [SID08] were selected. Notable information extracted from widely renowned Internet blogs such as [USA08] [FAC08] were also considered. Apart from these references, guidelines for current regulations on digital accessibility such as

UNE139801:2003 [UNE139801] and UNE139802:2003 [UNE139802] were mainly used. Table 3.5 lists the new heuristics for public kiosks.

Body expressions
Does the system express any emotion through any gesticulation?
What kind of emotions does the system show?
Are emotions related to the textual content?
Does the virtual assistant have a good behaviour according to the receivers of the information?
Does the virtual assistant include a logo or corporative image that represents the company?
Dialog
Does the system permit inserting questions in different languages?
Does the system have audio?
Does the system inform about the information that it has?
Are the messages of the system rotary, always different or they are changing constantly?
Are sentences coherent and do they suit the questions?
If the system does not have the needed information, does the system inform the user?
Can the user contextualize the question and the system presents the same results?
Is the system able to relate different questions to delimit the results of the search?
Does the system identify a change of the topic?
Others
Do the system responses arrive in a real and acceptable time?

Table 3.4 New heuristics for virtual assistant

Physical distribution
Is the system visible where it is located?
Are there indicators of the existence of the system nearthe system?
Is it possible to access the system with a wheelchair?
Is the location of the system not close to a highly trafficked area?
Are external connections for external devices easy to access?
Is the printer easy to access?
Software of device
Does the system use in all cases the same operating system?
Does the system use in all cases the same navigator?
Does the system have a screen reader for blind people or people with limited visibility?
Does the software permit users to customize the interface according to the standard configuration of the navigator?

Interaction
Does the height of the system permit independent interaction for all those who use a wheelchair or a very tall or very short person?
Is the system easy to use or does it require much effort?
Hardware (Screen)
Are the screen inclination and the reflected light adequate?
Is it possible to adapt the colour, the brightness and the contrast to the environmental conditions?
Does the screen offer a device to change the screen position?
Does the screen not blink?
Hardware (Mouse)
Is the mouse conventional?
If the mouse has additional buttons, is easy to know the goals of these additional buttons?
Hardware (Keyboard)
Do you not have to make additional efforts to push the keys?
Does the keystroke have a tactile sensation and/or a sonorous sensation?
Do the keys have a tactile mark (Braille language)?
Can you touch the keys without activating them?
Is colour not the only option to differentiate the keys?
Does each different group of keys use a different colour?
Are the labels of the keys readable?
Assistive hardware
If the system has assistive aids, does the system include the needed hardware to use these assistive aids?
Hardware (Audio)
Does the system emit sounds?
If the system emits some sound, is the audio information shown in a visual way on the screen?
Is the volume suitable?
Can the user modify the volume?
Does the system emit sounds?
Indicators
Is the indicator of the printer easy to locate?
Is the indicator of the digital certificate and electronic card easily visible?
Help and documentation
Does the system show help messages for its use?
If assistive posters or documentation exists, are these understood?
Is there an extra resource where the user can obtain help to use the interactive system?

Table 3.5 New heuristics for public kiosks

Therefore, the set of reviewed heuristics includes those for websites, desktop applications, virtual assistants and public kiosks. In total, 363 heuristics have been collected.

3.2.2 UX facets and ISO/IEC 25010:2011 attributes

Up to now, an exhaustive state of the art concerning usability heuristics definitions has been presented, but some research questions appeared: is usability the only needed or existing facet considered in the UX development process? Are there heuristics for other UX facets? What exactly does UX include? What relation exists between UX definition and the standard related to the product quality? In this section the answers to these questions are presented.

3.2.2.1 Concepts Involved in the UX

The meaning of most terms used in the UX field is not clear. Everyone uses what they prefer. Properties, facets, dimensions, features, sub-features, categories and attributes are concepts frequently used in the UX field and in the HCI discipline. However, what are the most appropriate expressions to call the concepts that each term refers to? To choose the best term for each concept, the definition of each term was looked up in two dictionaries. The Oxford dictionary was contrasted with the definitions from The Cambridge dictionary. Table 3.6 shows definitions from both dictionaries.

Concept	Oxford dictionary	Cambridge dictionary
Property	An attribute, quality, or characteristic of something.	An object or objects that belong to someone
Facet	A particular aspect or feature of something.	One part of a subject, situation, etc. that has many parts
Dimension	A measurable extent of a particular kind, such as length, breadth, depth, or height.	A measurement of something in a particular direction, especially its height, length, or width
Feature	A distinctive attribute or aspect of something.	A typical quality or an important part of something
Category	A class or division of people or things regarded as having particular shared characteristics.	A type, or a group of things having some features that are the same
Attribute	A quality or feature regarded as a characteristic or inherent part of someone or something.	A quality or characteristic that someone or something has

Table 3.6 Comparison of UX definition terms

The definitions of both dictionaries are very similar. The terms property, feature and attribute can all be considered as synonymous. In reference to the term dimension, it has the nuance of measurable.

> Therefore, the following terms will be used:
>
> **Facet** for determining all UX components. **Attributes** for all features, subfeatures and attributes in the standard, and, finally, the term **dimension** will be used for measurable quality attributes considered in ISO 2502n.

3.2.2.2 UX specification through quality attributes

On the one hand, the facets considered in UX are still not agreed on in the scientific community or in any type of organization for standardization. Related works in this area are [HAS06] [MOR05]. In addition, there are other concepts that UX can comprise: accessibility [W3C08], emotional [LEI08], communicability [PRA00], cross-cultural [JIA09], plasticity [THE99], playability [GON09] and reliable [AVI04], among others. Thus, one or another facet is used according to the author and their needs in the design or evaluation process.

On the other hand, SQUARE (Systems and Software Quality Requirements and Evaluation) (ISO/IEC 25010:2011) [ISO25010] is the term that refers to the standard that defines the system and software quality. Nevertheless, a relation between the standard attributes and facets that the community uses to evaluate the UX does not exist. Therefore, the specification of UX based on the facets that the ISO/IEC 2010:2011 considers due to its attributes is presented here.

Two groups of data are needed to get the UX facets that are implicitly included in ISO/IEC 25010:2011 [ISO25010]. The first group is the attributes of the standard. According to ISO, 42 standard attributes are considered in the research. The second group of data is the UX facets. This research is based on the facets mentioned in section 2.3.1. Therefore, 13 UX facets are considered in the analysis. Bearing in mind both groups of data, we are looking for the UX facets that are implicitly considered in each standard attribute. So, for each one of the 42 attributes, we have reviewed its definition and compared it to each of the 13 UX facet definitions to decide if the specific UX facet was implicitly considered in the standard attribute.

Finally, according to the specific definition of each facet and the specific definition that the ISO/IEC 25010:2011 presents for each attribute, facets which are implicitly

considered in the 42 attributes of the standard ISO are presented in Table 3.7 and Table 3.8.

		ISO/IEC 25010 : 2011	UX facets
4.1 Quality in use		**4.1.1 Effectiveness**	Usability, Playability, Useful
		4.1.2 Efficiency	Usability, Playability
	4.1.3 Satisfaction	4.1.3.1 Usefulness	Useful
		4.1.3.2 Trust	Emotional, Playability, Desirable
		4.1.3.3 Pleasure	Emotional, Playability, Desirable
		4.1.3.4 Comfort	Emotional, Playability, Desirable
	4.1.4 Freedom from risk	4.1.4.1 Economic risk mitigation	Dependability
		4.1.4.2 Health and safety risk mitigation	Dependability
		4.1.4.3 Environmental risk mitigation	Dependability
	4.1.5 Context coverage	4.1.5.1 Context completeness	Usability
		4.1.5.2 Flexibility	Usability and Accessibility

Table 3.7 UX facets for each attribute of Quality in use

	ISO/IEC 25010 : 2011		UX facets
4.2 Product quality	**4.2.1 Functional suitability**	4.2.1.1 Functional completeness	Useful
		4.2.1.2 Functional correctness	Accessibility, Playability
		4.2.1.3 Functional appropriateness	Accessibility, Playability
	4.2.2 Performance efficiency	4.2.2.1 Time behavior	Usability
		4.2.2.2 Resource utilization	Dependability, Accessibility
		4.2.2.3 Capacity	Dependability
	4.2.3 Compatibility	4.2.3.1 Co-existence	Plasticity
		4.2.3.1 Interoperability	Accessibility, Plasticity
	4.2.4 Usability	4.2.4.1 Appropriateness recognizability	Usability, Findable
		4.2.4.2 Learnability	Usability, Playability
		4.2.4.3 Operability	Usability
		4.2.4.4 User error protection	Usability, Playability
		4.2.4.5 User interface aesthetics	Usability, Playability
		4.2.4.6 Accessibility	Accessibility
	4.2.5 Reliability	4.2.5.1 Maturity	Dependability
		4.2.5.2 Availability	Dependability, Accessibility
		4.2.5.3 Fault tolerance	Dependability
		4.2.5.4 Recoverability	Dependability
	4.2.6 Security	4.2.6.1 Confidentiality	Dependability
		4.2.6.2 Integrity	Dependability
		4.2.6.3 Non-repudiation	Dependability
		4.2.6.4 Accountability	Dependability
		4.2.6.5 Authenticity	Dependability
	4.2.7 Maintainability	4.2.7.1 Modularity	Dependability
		4.2.7.2 Reusability	Dependability
		4.2.7.3 Analyzability	Dependability
		4.2.7.4 Modifiability	Dependability, Accessibility
		4.2.7.5 Testability	Dependability
	4.2.8 Portability	4.2.8.1 Adaptability	Accessibility, Plasticity
		4.2.8.2 Installability	Plasticity
		4.2.8.3 Replaceability	Plasticity

Table 3.8 UX facets for each attribute of product quality

3.2.2.2.1 Discussion

According to the research carried out about this issue, facets implicitly included in ISO/IEC 25010:2011 are *accessibility, dependability, desirable, emotional, findable, playability, plasticity, usability and useful*. Therefore, this set of facets should be the minimum for considering the ISO/IEC 25010/2011 in the design or evaluation of interactive systems.

Despite the results of this research and as UX experts, two more facets are needed in the design or evaluation process and when other facets are applied. These facets could work in a transverse way and they are called communicability [PRA00] and cross-cultural [JIA09]. Even though the standard does not explicitly present attributes about communicability and cross-cultural, it is necessary to consider both facets in any part of the design or evaluation process due to the importance of the internationalization to reach more popularity. Therefore, these facets should be considered at the same time or in the same way as the other considered facets. In this context, two types of facets are differentiated.

- The parallel facets (dependability, usability, playability, plasticity, accessibility, emotional, desirable, findable and useful) can be applied in an interactive system by themselves (in an individual way).
- Transverse facets (communicability and cross-cultural) can be applied at the same time as when another facet is applied.

Figure 3.1 represents this definition.

In reference to the analysis performed, it is important to highlight that there are attributes that are considered in more than one UX facet. Therefore, the non-isolation of UX facets is consolidated. In addition, all standard attributes are considered by some UX facets; in fact it causes a direct relation between both facets and standard attributes.

The number of attributes detected in each facet is presented in Table 3.9. Even though usability is the most used facet to be considered when designing or evaluating an interactive system, dependability wins the first position including 19 attributes for the standard. Usability is in second position with 10 attributes, the same number of attributes as playability. The fourth position is for accessibility,

which includes 9 ISO attributes. And then, plasticity with 5 and emotional, desirable and useful with 3 attributes. The last position is for the findable facet with only one attribute.

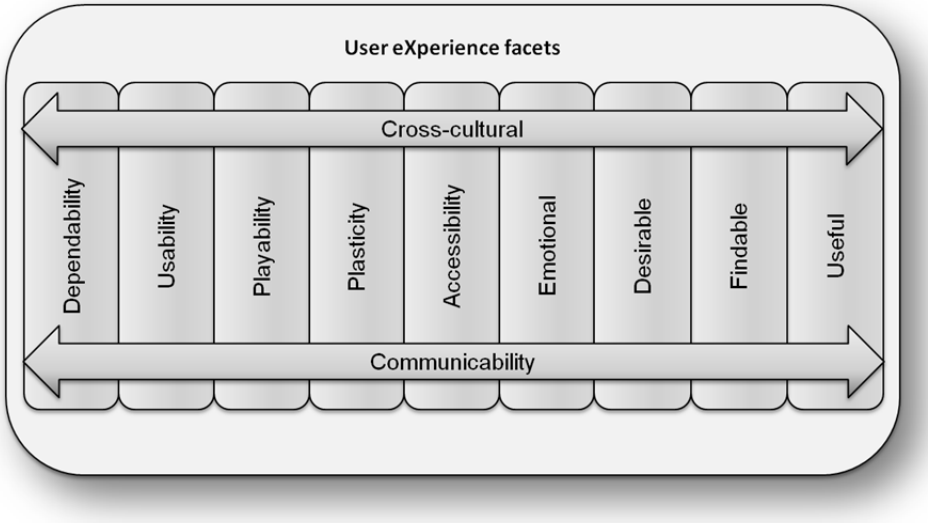


Figure 3.1 UX facets

We consider that usability occupies the second position and not the first due to the tendency of including in the usability facet several features that are more suitable in other ones. Therefore, if usability only considers the exclusive properties of usability, the number of ISO attributes that usability includes decrease and it causes a rise in the number of attributes of other facets as is the case with dependability.

UX facets	Number of attributes
Dependability	19
Usability	10
Playability	10
Accessibility	9
Plasticity	5
Emotional	3
Desirable	3
Useful	3
Findable	1

Table 3.9 Number of attributes in each UX facet

Up to this point and as we detailed in Chapter 2, different definitions concerning UX exist, but depending on the authors or the specific needs for the design or

evaluation, some UX facets are used taking into account which facets are more adequate for the specific interactive system.

Then, our proposal includes a more general definition of UX, which covers all the aspects commented on previously:

> "User experience considers all aspects, internal as well as external of the user and interactive systems, which provoke any feeling in whoever uses the interactive system in a specific context of use."

3.2.2.3 Heuristics classification in different UX facets

Before describing the specific process for detecting specific facets for each standard attribute, it is essential to describe the certain state of the heuristics and facets. Up to now, this project has talked about usability heuristics but different UX facets are detected.

As I have previously pointed out, usability is the oldest and most often used facet. Therefore, the HCI community tended to put all types of heuristics in the same usability bag. Due to this comfortable and usual trend, usability increased and increased covering any kind of heuristics. To cut this trend, the heuristics collected up to now were analysed to put each one in the most appropriate UX facet. Currently there are 363 heuristics distributed in each specific UX facet as shown in Table 3.10. **Appendix B** lists an example of heuristics for the designer of "Communicability" facet.

Facet	Number of heuristics
Reliability	27
Usability	239
Accessibility	15
Emotional	6
Findable	19
Communicability	41
Cross-cultural	16

Table 3.10 Number of heuristics in each UX facet

3.2.2.4 New heuristics for ISO/IEC 25010:2011

Up to now, the UX facets that are included in each attribute of the ISO/IEC 25010:2011 standard are presented. However, this is not enough to get the set of heuristics for a specific interactive system that is the most suitable if the main goal is to apply the ISO standard. Therefore, the most appropriate heuristics for each facet of each attribute should be detected.

In this section the process to get the heuristics for each attribute of the standard is specified. The process includes three main steps: (i) reviewing the heuristics collected up to now. (ii) Reviewing the heuristics that were not considered in the first review and (iii) creating new heuristics for the attributes that do not have any heuristics.

3.2.2.4.1 Reviewing the heuristics of the initial review

Before creating new heuristics for some attributes of the standard that will be detailed later, the initial review presents 363 heuristics that are divided into different UX facets as Table 3.10 shows.

Therefore, according to the UX facets detected in each attribute of the standard, each heuristic definition was valued in each attribute definition to decide if the heuristic is included in this attribute or in another one.

As a result, one heuristic can be considered in more than one attribute of the standard. Therefore, the whole set of 363 heuristics are included in at least one attribute of the standard. However, some ISO attributes do not accept or receive any heuristics, so some attributes have heuristics but others do not. **Appendix B** details an example of heuristics for the designer of "Recoverability" attribute of the standard.

3.2.2.4.2 Reviewing the heuristics that were not considered in the first selection of heuristics

When the first review of heuristics from 1986 to 2010 was carried out, some heuristics were not considered (see [MAS10a] for reasoning and details). At this point, for each attribute of the standard that does not include any heuristic, a review of the previous heuristics not considered was done to check if some of these heuristics not considered could be included in some of these attributes.

At this point, the heuristics not considered that can now be considered in the ISO attributes are presented in Table 3.11.

ISO attribute	Heuristics
Health and safety risk mitigation	Is the system adapted to the context conditions? (the category of Context by Constantine [CON95])
Functional correctness	Is the technology used transparent for the user? (the Invisible Technology by Mayhew [MAY92])
Functional appropriateness	Is it possible to reject default values in a fast and easy way? Are the expressions "by default", "standard", "using usual values" or "reestablish the initial values" avoided? (the Default Value principle by Tognazzini [TOG03])

Table 3.11 New heuristics from not considered ones

ISO attribute	Heuristics
Pleasure	• After using the system, are your needs satisfied?
Comfort	• Are you comfortable using the interface? Are you helped by the interface?
Economic risk mitigation	• Does the system avoid mitigating the potential risk to financial status in the intended context of use? • Does the system avoid mitigating the potential risk to efficient operation in the intended context of use? • Does the system avoid mitigating the potential risk to commercial property in the intended context of use? • Does the system avoid mitigating the potential risk to reputation in the intended context of use? • Does the system avoid mitigating the potential risk to other resources in the intended context of use?
Health and safety risk mitigation	• Does the system or the product avoid mitigating the potential risk to people in the intended contexts of use?
Functional completeness and usefulness	• Does the set of functions cover the entire specified task? Does the set of functions cover all the user objectives?
Functional correctness	• Does the system provide the correct results with the needed degree of precision? • Do the functions facilitate the accomplishment of specified tasks and objectives?
Resource utilization	• Is the system using the needed CPU resources when performing its functions?
Co-existence	• Can the product perform its required functions efficiently while sharing a common environment and resources with other products?
Authenticity	• Is the identity of the subject or resource the one claimed? • Has the interface a copyright? • Is the information original and not copied from another source?
Modularity	• Is the system or the computer program composed of discrete components? Is it possible to change one component with the minimal impact on other components?
Reusability	• Can an asset be used in more than one system? • Can an asset be used building other assets?
Installability	• Can the system or the product be successfully installed and/or uninstalled in a specified environment?
Replaceability	• Can the system or the product be upgraded easily?

Table 3.12 New heuristics

3.2.2.4.3 Creating new heuristics

Even though a review of heuristics was carried out, some attributes of the standard still do not have heuristics.

Therefore, we think it is essential to create new basic heuristics to cover the attributes that, according to our research, do not present any other heuristic.

Bearing in mind the brief definition of each attribute presented in the ISO and the UX facet definition where the attribute is included, some new heuristics were defined. The complete list of new heuristics is detailed in **¡Error! No se encuentra el origen de la referencia.**.

Most attributes have different heuristics but two attributes still remain without any heuristics. These attributes are analyzability and testability. Following the approach of this research, we would like to obtain heuristics to consider UX in the design and evaluation process. Both attributes refer to the capacity that the interactive system has to be analysed or tested. Therefore, considering that our approach is already detailed in the evaluation and design step of the development process, we consider that both attributes, analyzability and testability, cannot be considered for improving the UX of an interactive system.

3.2.3 Components

At this point, a large set of heuristics is detailed and the UX facets and standard attributes are also selected. This information is considered in the repository of information through a methodology based on heuristics for considering UX in the design or evaluation stage of the development process of interactive systems. However, there is more needed information.

This section is for the description of system components as needed information in the repository.

Components of interactive systems are all parts of an interactive system, as well as software parts and hardware parts.

The origin of this section is when the case studies were started. I considered that it is important to have any kind of limit to validate that the set of collected usability heuristics covers as many as possible usability features of the interactive system.

For this reason, I tried to divide the interactive systems of the case studies (public kiosk and virtual assistants) into a limited list of "things" that permits the selection of the most suitable heuristics for each specific case.

Therefore, I spent time to make some efforts to detect public kiosk and virtual assistant components. And, obviously, common websites were also considered due to it being one of the most usual interactive systems. Then, a classification of interactive system components was achieved. It allows me to know what parts of the interactive systems should be assessed and to choose the best heuristics for each part.

As a result, components were divided into two big groups: software and hardware interactive system components. This classification is presented because software and hardware systems have different features or components that involve different types of specific heuristics. Then, some general categories of components are divided into more specific components to allow a complete classification of interactive system components in order to facilitate their evaluation.

First of all, an exhaustive search in literature to find any classification of interactive system components was done but, unfortunately, I did not find many researches about it.

Elements detected by Jakob Nielsen and Kara Pernice in their new book "Eyetracking web usability" [NIE10] were found. The book presents fundamental website design elements. Their list of elements permits designers to consider all the parts in a website design but they do not specify enough to consider all of the elements in our classification.

Another study related to the classification of interactive system components is [GON10]. In this research, a classification of videogames parts is presented. The way that the author developed his idea is very interesting for achieving a conceptual model of the parts detected in this type of interactive system. However, this model only includes specific interactive systems and it does not cover a larger set of them.

Therefore, a classification that includes specific parts of the most used interactive system such as common websites, virtual assistants or public kiosks was not found. In the following, the classification of interactive system components including all parts of these interfaces is detailed. The hierarchy saves 57 components that are divided into software and hardware components. Table 3.12 shows the number of heuristics that both components have.

Components	Number of heuristics
Software	442
Hardware	85

Table 3.12 Number of heuristics for each component

3.2.3.1 Software components

Software interactive systems are the systems that include some type of software such as websites, desktop applications, video recorders, or mobile phone applications among others. But in addition, all hardware systems that also include software parts (although this software part will be a little part of a big hardware system) should also be considered. For instance, public kiosks or some electrical appliances, and others, have important hardware components but they also provide a little software interface that cannot be forgotten when a complete UX evaluation is carried out.

According to common websites, virtual assistants and desktop applications, some components that are susceptible to UX evaluation can be detected. Therefore, the classification of software components shows a set of components that can be detected in many types of interactive systems. To sum up, Table 3.13 presents a classification of software interactive system components that can be used to detect each part of most software interactive systems.

General group	Specific component
Type of content	Forms, tables, lists, dates, times, numerical values, money signs
Information	Pictures, news, graphics, format, text, URL, abbreviations, audio, nomenclature, colors, icons
Data management	Information transmission, sign in form, log in form, updating information, information validation, information recovery
Search	Search form, results of search
Navigation area	Pages, titles, cursor, shortcuts
Emergency exits	--

Table 3.13 Classification of software components

3.2.3.2 Hardware components

Hardware interfaces are known as some physical device that you can interact with. Some examples are elevator panels, public kiosks, video recorders, ovens...

As above, the challenge was choosing which “hard” component of the public kiosk should be analysed. A public kiosk, as a physical system, is a complete interactive system with a large set of components and a little software part.

So, hardware interfaces are classified into operative systems, browser, in/out devices, assistive aids, audio, indicators and help and documentation. Then, every category is divided into more components to allow a complete hardware component classification in order to facilitate its evaluation.

The hardware components categorization is detailed in Table 3.14.

General group	Specific component
Operating system	--
Browser	--
In/out devices	Printer, digital certificate, electronic card, mouse, screen, keyboard
Assistive aids	Icons, assistive hardware
Audio	Volume
Indicators	--
Help and documentation	--

Table 3.14 Classification of hardware components

3.2.3.3 Discussion

Note that hardware classification includes software components such as operating systems or browsers. It is because hardware systems might provide software applications and we consider that the same type of hardware interactive system should include the same version of operating system or browser to avoid confusion for users when they use these systems.

In addition, hardware systems also include software components. Therefore, this classification is not a mutually exclusive classification but it is a complementary classification because the hardware and software components of systems might be found in a particular interactive system such as public kiosks.

It is also obvious that sometimes some heuristics can take part in more than one part of an interactive system. For example, there are some heuristics like “the same things have to be aesthetically equals” which are applicable in indicators, keys, icons, among others. Therefore, some heuristics are the best for more than one of the interactive system components. However, if you know perfectly which components of the interactive system you want to evaluate or which components should be considered, you will be able to choose the best heuristics for each component easier.

3.2.4 Functionalities

```
Functionalities are the description of all tasks that
might be carried out through the interactive system and
where the user should have a positive experience.
```

Functionality implies a different set of heuristics according to their own nature. Table 3.15 shows the set of functionalities considered in this project and the number of heuristics for each one.

Functionality	Number of heuristics
To look up information	129
To do procedures	232
To register	18
To log in	16
To fill in forms	72
To print conformations	4

Table 3.15 Number of heuristics for each functionality

3.2.5 Features

```
Features are all these aspects, that are not
functionalities, and they should be considered when the
UX is included in the design or evaluation of an
interactive system.
```

For example, the virtual assistant should simulate human behaviour. Therefore, this is a feature that implies the inclusion of emotional heuristics and heuristics of gesticulations. The number of heuristics for each considered feature is presented in Table 3.16.

Feature	Number of heuristics
It has a software part	294
It has a hardware part	40
It simulates the human behavior	18

Table 3.16 Number of heuristics for each feature

3.2.6 Interactive systems

> Interactive system is defined as any type of system with any type of (software or/and hardware) interface to interact with it.

Therefore, an interactive system is composed of components, functionalities and features and, as we have seen before, there is heuristics for components, functionalities and features. This means that it will be able to easily provide the heuristics concerned with any interactive system (to be designed or evaluated).

Due to all semantics related with interactive system concepts, interrelations among them are presented through ontology in the following section.

3.2.7 Ontology of the repository of information

The previous 6 sections detailed the information needed to optimize the heuristic evaluation methodology. Now, descriptions, relations and semantic meanings are presented through ontology that defines an interactive system and represents the needed semantic meaning to enable the selection of the most suitable heuristics for every specific case. Therefore, this section details the description of components, functionalities, features, facets, standard attributes, heuristics and interactive systems and its semantic relations. The tool used for constructing the ontology was Protégé (http://protege.stanford.edu/) which designs ontologies with OWL language and provides a useful interface for saving figures such as the ones presented in the following sections.

Functionalities and components are defined using the field name and functionality or component parent, respectively. It means that functionalities and components can present a hierarchy. Figure 3.2 shows the fields needed to define functionalities and components.

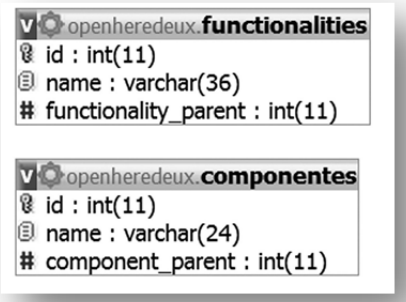


Figure 3.2 Functionality and component definition

As shown in Figure 3.3 and Figure 3.4, features are defined with a name and facets using a name and a description.

Figure 3.3 Feature definition

Figure 3.4 Facet definition

Furthermore, as it is presented in Figure 3.5, attributes of the standard ISO are defined with a name, a description and a parent because of the hierarchy presented in the standard.

Figure 3.5 Attribute definition

Finally, heuristics are defined through an evaluator description (as a question), a designer description (as a declarative sentence) and a heuristic parent to allow the creation of categories of heuristics. Figure 3.6 shows the definition.

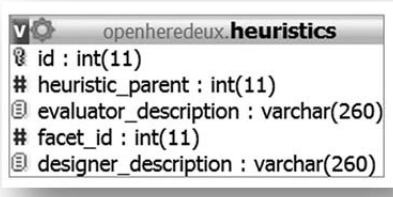


Figure 3.6 Heuristic definition

All this information, components, facets, functionalities and features, is related to heuristics through the relation "*HasHeuristic*". Due to the exhaustive research to determine the validated set of UX facets and components, heuristics are included in at least one facet and one component. However, not all heuristics are part of some functionality or feature. Specific research is needed to determine a group of functionalities and features that are considered in interactive systems.

Then, the semantic relation called "*HasHeuristic*" is represented in Figure 3.7. This is a key relation because of the presentation of the semantic meaning needed for determining the suitability of one heuristic for a specific functionality, feature, component, UX facet and attribute of the standard.

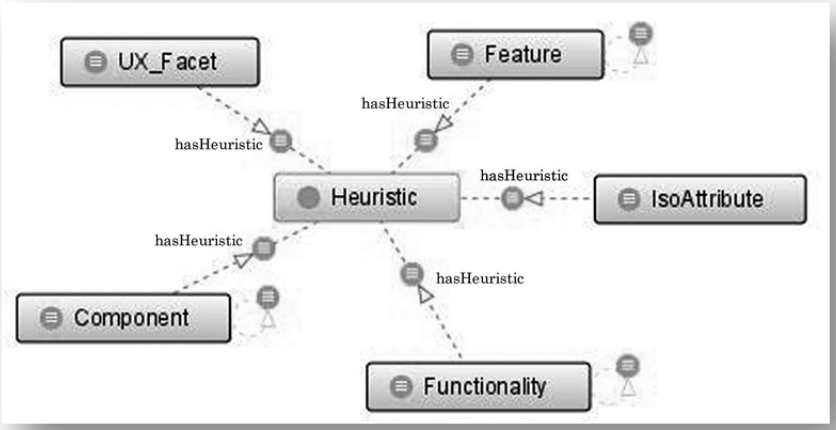


Figure 3.7 HasHeuristic semantic relation

Now, the specific definition of the interactive system is presented. Interactive system is defined as a name and a description as shown in Figure 3.8.

Figure 3.8 Interactive system definition

According to ontology, the key relations for acquiring the appropriate semantics are the relations “hasFunctionality”, “hasFeature”, “hasComponent” and “hasFacet”. In Figure 3.9 these semantics are presented.

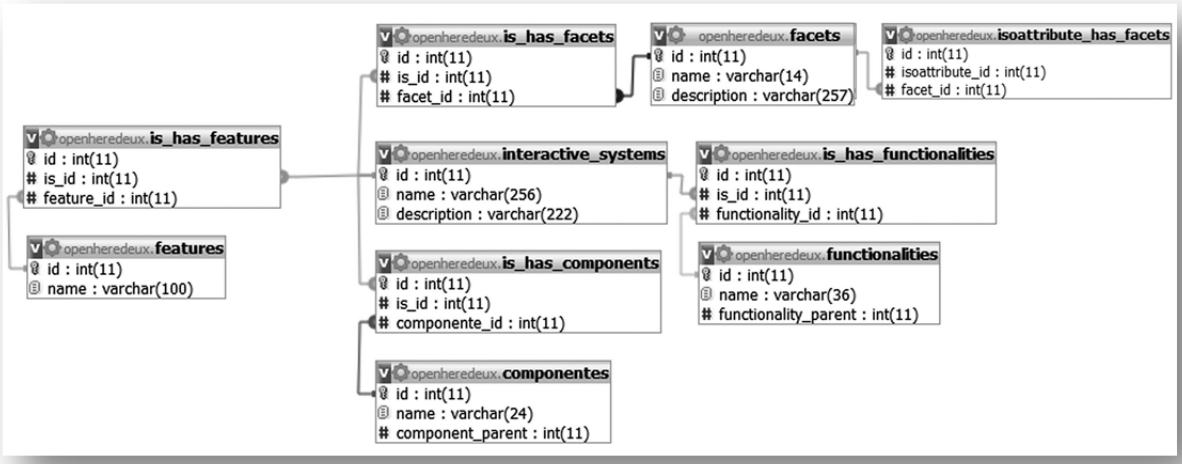


Figure 3.9 Semantic relations for the interactive system definition

3.3 Heuristic suggestion

As we have seen, the second thing lacking in heuristic evaluation methodology is how to answer this question: "*What is the most appropriate set of heuristics for a given interactive system (to be designed or to be evaluated)?*"

Once a huge set of heuristics is collected, the most suitable ones should be selected according to the main goals of the design or evaluation, understanding that the suggestion can be used in the design or evaluation step of the development process.

Nowadays, this process is a manual process and it hardly depends on the professional experience of the one preparing the evaluation. The best option for covering this inadequacy will come with an automatic procedure that provides a list of the most appropriate heuristics for a specific project (which has specific goals to be considered in the design or evaluation step).

In the following sections, the pre-process of suggesting the most suitable heuristics is detailed. First of all, the definition of specific goals of the suggestion is proposed. Then, two different financial constraints (UX degree and time) are presented in section 3.3.3. These financial constraints filter the number of suggested heuristics to be able to adapt this task of the development process to the pre-established budget. Afterwards, how to document possible conflicts among heuristics is proposed using a rationale notation called TEAM notation. Finally, the algorithm of suggestion is

presented through two different views: the process for the end user and the internal process.

3.3.1 Goals of the design or evaluation

Before getting the most suitable set of heuristics for a specific interactive system, the main goals should be as clear as possible. This means that when UX is considered in the development process, some aims appear in your mind. For example, the interactive system presents a new component such as a printer. Or maybe the system offers the functionality of shopping but nobody uses it. It is also possible that you would like to consider the communicability of the whole system. All these types of refinements are goals to be considered in the automation of the selection of the most suitable heuristics for a specific interactive system.

```
Therefore, according to the needed information and the
ontology presented previously, there are different
goals: the ISO/IEC 25010:2011 attributes, the aspects of
interactive system (such as functionalities, components
and features) and UX facets.
```

This means that the first step of the suggestion process is the choice of the main goals such as some or complete attributes of the standard, specific functionalities, features or components or different UX facets to apply in the interactive system.

3.3.2 Types of heuristics according to the receiver

Regarding the validation of the set of heuristics, the first hypothesis in reference to the writing of the heuristics was that heuristics should be written as declarative sentences for designers and as questions for evaluators.

After performing the validation, due to the evaluators' support of this hypothesis, two different types of suggestions are proposed.

On the one hand, heuristics for designers are suggested as declarative sentences to be design recommendations that designers can apply in the design of a specific interactive system or improvements in the extraction of qualitative results.

On the other hand, heuristics for evaluators are proposed as questions to be answered in the evaluation process. Therefore, the way to score the heuristics for a

specific interactive system is easier than if heuristics are presented as a declarative sentence. Table 3.17 presents an example of both types of suggestions.

Heuristics for the evaluator	Heuristics for the designer
Cross-cultural - Do updated dates and times appear in essential cases as prices? - Do the last time and date that the information is updated appears? - Does the time zone appear? - Are a.m and p.m abbreviations used? - Is the month written or abbreviated but not in a numerical form?	**Cross-cultural** - Update dates and time should appear in essential cases as prices - The last time and date that the information is updated should appear - The time zone should appear - a.m and p.m abbreviations should be used - The month should be written or abbreviated but not in a numerical form

Table 3.17 Example of heuristics for designers and for evaluators for the case of Cross-cultural facet

3.3.3 Financial constraints

As is well-known, the budget of a project is the most important factor to be strictly considered in companies. When the estimation of the needed resources is defined, it should be firmly applied in every task of the project. Therefore, the budget adjustment is an inescapable rule that project managers should apply.

Moreover, in a tech project, usually new appearances might increase the initial budget. Appearances such as new requirements when the product is already developed, delays in serial tasks of the project or technologic constraints cause the increase of the needed time to finish the product and in consequence, the increase of the initial budget.

The development process of an interactive system through a user centered design methodology [ABR04] is the type of project that could use these financial constraints. It includes six main steps that should be followed in an iterative way: requirements analysis, design, prototyping, implementation, evaluation and commercial launch [GRA03]. Therefore and in the same case as other kinds of projects, there are financial constraints in every part of the development process.

The financial constraints proposed here focus on the design and evaluation of an interactive system. Bearing in mind the defined process to carry out heuristic evaluation, the research questions are: “How much time does the project have to spend in the heuristic evaluation? Then, if the project has a kind of time restriction,

which set of heuristics should be considered? Are some more important than others?"

> In order to answer the research questions, the User eXperience Degree (UXD) and the time restriction are proposed as financial constraints that will be applied to the heuristic evaluation.

There are few projects where scientists carried out strenuous efforts to classify heuristics in different levels of importance. In [SPI12] the heuristics are classified in four levels: very high, high, medium and low. In another research, the heuristics are classified [USD06] through the "relative importance" of the guideline.

In the next section the definitions of the UX degree and the time restrictions are presented. Then, the survey to reach the UX degree for websites application is detailed.

3.3.3.1 UX degree

The main goal of the UX degree is the classification of the whole set of heuristics in more accurate sets to be able to not consider some of them in the case of budget restrictions. It presents different levels of consideration according to the importance that a specific heuristic has in a specific kind of interactive system.

> Therefore, the UX degree represents the level of importance that every heuristic (or full set of heuristics) has in a specific system.

Furthermore, and following accessibility levels [W3C08], three UX degrees are proposed:

> **U degree:** the heuristics of the U degree are essential to assure that the user who will use the interactive system will get a positive experience.
>
> **UU degree:** the heuristics of the UU degree are necessary to assure that the user who will use the interactive system will get a positive experience.
>
> **UUU degree:** it is advisable to consider the heuristics of the UUU degree to assure that the user who will use the interactive system will get a positive experience.

Following the accessibility guidelines example, heuristics belonging to the U degree are the minimum necessary to consider that those who use the evaluated interactive system will feel a little bit of a positive experience. However, if the interactive system should present a higher level of quality, it would be necessary to consider the three levels (U, UU, UUU).

3.3.3.2 Time per heuristic

Everyone agrees that "Time is money" ("*el temps és or*" in Catalan), in the business context.

This research also faces the challenge of limiting the number of recommended heuristics considering the time needed for the evaluation.

The time parameter is not obvious, it depends on the type of interactive system, the experience of the professional in the usage of the inspection methodology and the knowledge that the professionals have about the heuristics (if they are familiar or not with the set of the suggested heuristics). However, real industrial projects always need to estimate how much time is needed for everything, and UX evaluations do not escape from this consideration. Therefore, if it is possible to provide project managers with this information, they will be able to deliver a set of heuristics according to the specific budget of this task of the project.

In the following, we detail how the values of the UX degree and the time restriction are reached.

3.3.3.3 Setting values for the UX degree and time restrictions

The process to determine the values for the UX degree of some interactive system and the time restrictions is carried out through a survey. The first goal of the survey is to define a UX degree for the heuristics that can be applied in a specific interactive system: website applications. The second goal is to determine the approximate needed time to consider every single heuristic in this specific interactive system.

But first of all, the details about the set of heuristics used in the experiment are presented. Then, the non-confidential aspects about the participants are given to continue with the process followed to set values for UX degree and time. Finally, the results are detailed.

3.3.3.3.1 The set of heuristics

Due to our approach where UX includes different facets, the heuristics for websites are categorized in different facets. In Table 3.18, the number of heuristics for each

UX facet is shown. In total, a set of 267 heuristics is considered to evaluate UX for a website application.

UX facet	Number of heuristics
Cross-cultural	12
Communicability	29
Findable	14
Accessibility	2
Dependability	23
Usability	187

Table 3.18 Number of heuristics applicable in website applications

3.3.3.3.2 Participants

Participants in the survey should be HCI practitioners. Therefore, our option to recruit participants was to send a "Call for participation in a PhD research" to HCI experts, mainly university HCI researchers and UX professionals from different international companies. The main aspects of the "Call for participation in a PhD research" are:

- The contextualization and the main aims of our project.
- The specific interactive system that the participant should use to answer the survey (general websites).
- The definition of the UX degree (in the same way that it is presented here).
- The steps to answer the two questionnaires (they will be detailed in the next section).

In Table 3.19 and Table 3.20, the user profile is presented. The call for participation was sent to 79 HCI/UX experts from whom I obtained 63 answers (30 males and 33 females).

Age	Number of users
18-25	3
26-35	33
36-45	15
46-56	12

Table 3.19 Age of the users

Years of experience	Number of users
Less than 2	3
2-5	30
6-10	24
More than 10 years	6

Table 3.20 Years of experience

3.3.3.3.3 Process

Due to there not being an organization regulating UX, the most convenient way to determine if one heuristic has one or another UX degree is by asking experts. Therefore, I asked for the UX degree of the set of heuristics that are applicable in website applications. The process that each participant followed to participate in the survey was divided in three main steps:

- Fill in the user profile form.
- Answer the first questionnaire selecting the degree that the participant considers more suitable for each heuristic taking into account that the set of heuristics is for a website.
- Answer the second questionnaire with questions about the time.

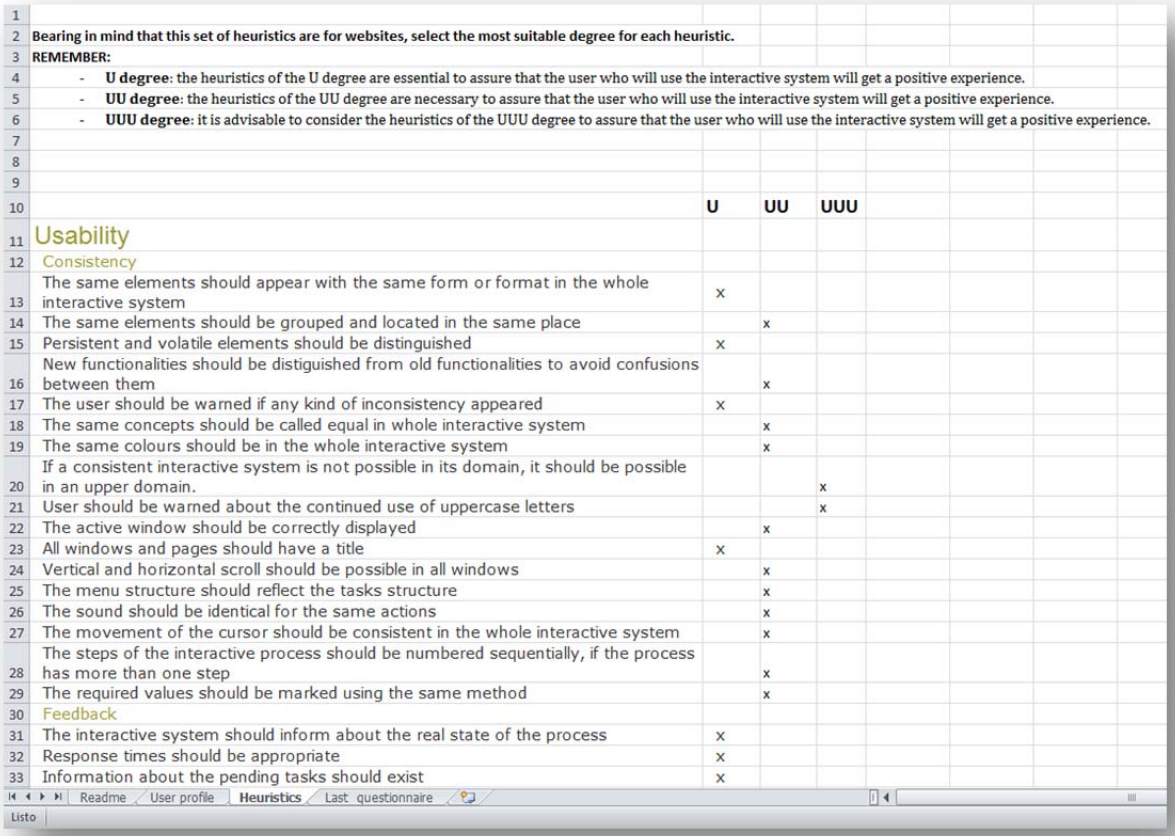
The first part presented the introduction of the survey including the main aim of the research, the process to follow and some other information needed to answer it. This information was the same as the "Call of participation in a PhD research" email.

The second part searched the user profile information. According to the information presented in the previous section, questions about gender, age and experience in the HCI field were proposed.

In the third part, each participant determined the UX degree for each heuristic through a questionnaire. Figure 3.10 shows one example of the questionnaire answered by one participant. Finally, the fourth and last part presented another questionnaire where some aspects such as time restrictions were asked.

Taking into account the large amount of heuristics, the set of heuristics was divided into 3 groups to provide the participants with a shorter group of heuristics. Thus, in the first group, the set (Q1) included heuristics from cross-cultural, communicability, findable, accessibility and dependability (80 heuristics). The second (Q2) and the third (Q3) groups included half the amount of usability heuristics (88 and 99 heuristics each one). The usability facet was divided into two

groups because the initial set of heuristics was so wide to inspire participants in the answering of the questionnaire.



1				
2	**Bearing in mind that this set of heuristics are for websites, select the most suitable degree for each heuristic.**			
3	**REMEMBER:**			
4	- **U degree**: the heuristics of the U degree are essential to assure that the user who will use the interactive system will get a positive experience.			
5	- **UU degree**: the heuristics of the UU degree are necessary to assure that the user who will use the interactive system will get a positive experience.			
6	- **UUU degree**: it is advisable to consider the heuristics of the UUU degree to assure that the user who will use the interactive system will get a positive experience.			
7				
8				
9				
10		**U**	**UU**	**UUU**
11	Usability			
12	Consistency			
13	The same elements should appear with the same form or format in the whole interactive system	x		
14	The same elements should be grouped and located in the same place		x	
15	Persistent and volatile elements should be distinguished	x		
16	New functionalities should be distiguished from old functionalities to avoid confusions between them		x	
17	The user should be warned if any kind of inconsistency appeared	x		
18	The same concepts should be called equal in whole interactive system		x	
19	The same colours should be in the whole interactive system		x	
20	If a consistent interactive system is not possible in its domain, it should be possible in an upper domain.			x
21	User should be warned about the continued use of uppercase letters			x
22	The active window should be correctly displayed		x	
23	All windows and pages should have a title	x		
24	Vertical and horizontal scroll should be possible in all windows		x	
25	The menu structure should reflect the tasks structure		x	
26	The sound should be identical for the same actions		x	
27	The movement of the cursor should be consistent in the whole interactive system		x	
28	The steps of the interactive process should be numbered sequentially, if the process has more than one step		x	
29	The required values should be marked using the same method		x	
30	Feedback			
31	The interactive system should inform about the real state of the process	x		
32	Response times should be appropriate	x		
33	Information about the pending tasks should exist	x		

Readme | User profile | **Heuristics** | Last questionnaire

Listo

Figure 3.10 Real questionnaire where the UX degree should be selected

Finally, the 3 different sets of heuristics were sent to HCI professionals via email and they had two weeks to fill in the questionnaire. Q1 was sent to 26 people. Q2 was sent to 30 people and Q3 was sent to 23 people. But, unfortunately, not everybody answered the questionnaire. Q1 was answered by 21 participants, Q2 by 18 and Q3 by 24.

Furthermore, the results were not directly reached using the answers of the participants. It was a more complicated process. In the first step, the UX degree of the most heuristics was defined according to the most selected UX degree. But in some cases it finished in a draw. This means that two or three different UX degrees

were voted by the same number of participants in a specific heuristic. For example, in the heuristic "The active window should be correctly displayed" from Q2, 9 participants marked it as U and 9 participants marked it as UU. To solve the draw, the heuristics in draw were sent again to participants who had answered the questionnaire previously. Therefore, the Q1-in-a-draw was sent to 21 participants and it presented only 12 heuristics, Q2-in-a-draw was sent to 18 people with 11 heuristics and Q3-in-a-draw was sent to 24 users including 9 heuristics. However, once again, not everybody answered the new and short questionnaire. Only 12 people answered Q1-in-a-draw, 12 people sent the Q2-in-a-draw and 24 participants answered the Q3-in-a-draw. Although this new round, 6 draws were repeated. We decided to consider the more strict degree in any case (U is more strict than UUU). In the next section, the results are detailed.

3.3.3.3.4 UX degree results

Obviously, the results of this research are the UX degree determined for each heuristic of websites applications and the time needed to spend in the evaluation of each heuristic. The number of heuristics that each facet has in each UX degree is presented in Table 3.21:

UX facet	UX degree		
	U	UU	UUU
Cross-cultural	6	5	1
Communicability	19	8	2
Findable	12	1	1
Accessibility	1	1	0
Dependability	15	7	1
Usability	91	75	21

Table 3.21 Number of heuristics for each facet and for each UX degree

3.3.3.3.5 Time restriction results

As I have pointed out in the previous section, the factor of the time restriction was asked in a second questionnaire. Two specific questions were asked:

How much time do you think that you need to score one of these heuristics during an evaluation? (question1)

Is one minute for evaluating two heuristics enough? (question2)

Bearing in mind those 63 experts that took part in the research, Table 3.22 and Table 3.23 presents the answers of both questions related to the time restrictions:

	Number of participants
Depends on the heuristics	9
Less than 20 seconds	9
20 seconds-1minute	27
More than 1 minute	15
No answer	3

Table 3.22 Answers for question1

	Number of participants
It is enough	42
It is not enough	9
Depends on the heuristics	9
No answer	3

Table 3.23 Answers for question2

According to question1, UX professionals and HCI researchers are not really sure about the time needed to score a heuristic. But in most cases participants think that a minute is enough to evaluate one heuristic. In reference to question2, 42 of 63 participants assure that is possible to score 2 heuristics in only one minute.

Therefore, the consideration of **one minute to evaluate two heuristics** is a good option for quantifying the amount of heuristics needed according to the budget of the task.

3.3.3.4 Discussion

Both financial constraints are focused on enhancing the UX evaluation in real cases. It enables a more accurate evaluation scheduling and a much better use of the project budget.

The definition of a UX degree for heuristics that are applicable in website applications is the first step to get a standardization of the UX evaluation. If a consolidation of these results is reached, the certification of the UX for different interactive systems could be possible.

In future work, we plan to define the UX degree for a wide variety of interactive systems: public kiosks, virtual assistants, mobile applications, etc.

3.3.4 Conflicting heuristics

The resolution of conflicting heuristics is a daunting and demanding task that often requires taking into account the trade-offs associated with alternative design choices. Therefore, whenever a good solution for solving conflicts among heuristics is found, it is worth the effort recording and documenting it for further reuse. As we have detailed in section 2.5.3, our proposal defines that conflicting heuristics are described with rationale notation called TEAM notation.

This part of the PhD was carried out with the collaboration of the Interactive Critical System research team from the Paul Sabatier University (Toulouse, France) where I had the chance of working with them for three months through a research stay. In a previous work, as detailed in section 2.5.3.1, they proposed TEAM notation [MAR10] as a rationale notation for making design decisions. Thus, the main goal of the collaboration was to define how to represent conflicts among heuristics using TEAM notation and facilitating the decision-making process.

Furthermore, it is important to highlight that the nomenclature used in the original research and publication about conflicting heuristics is not exactly the same as used in this PhD. Bearing in mind that I was working in their research group, we used their concepts. Now these concepts are used in this PhD. Then, for preserving the originality of this research Table 3.24 shows the equivalences that this PhD should consider taking the advantage of the research carried out in the stay.

ICS Team notation	This PhD notation
Guideline	Heuristic
Factor	UX facet
Criteria	Standard attribute
Database	Information repository

Table 3.24 Nomenclature equivalence

This section presents how a heuristic is defined using design rationale elements. Then, the types of conflicts and how they are described are presented. Finally, how to document these conflicting heuristics for further usages is detailed.

3.3.4.1 Mapping individual heuristic to design rationale elements

Heuristic sources often provide information that can easily be matched to UX facets and attributes of the standard ISO/IEC 25010:2011 as is illustrated in Figure 3.11. Hereafter an example based on the "WCAG guideline 1.1 text alternatives" is provided in the following:

- **Description**: "1.1 Text Alternatives: Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language".
- **Source**: (WCAG) 2.0 (see [W3C08])
- **UX facet**: Accessibility
- **Standard attribute**: Perceptibility

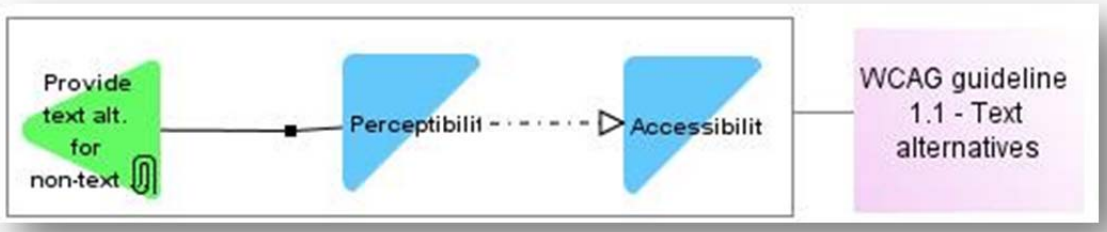


Figure 3.11 Representation of individual heuristic using the TEAM notation

It is noteworthy that the overall description of the heuristic is mapped to a non-functional requirement (represented by a square). Moreover, the heuristic is connected through a box embedding the criteria "provide text alt. for non-text", the UX facet accessibility and attribute perceptibility; such as composition shows that all these elements are part of the heuristic description. The clip in the diagram indicates that there are additional documents explaining how the design option can be assessed.

3.3.4.2 Conflict management and heuristics cleaning

As listed above not all conflicts look alike and many types of connections between heuristics might occur. For example, considering the existence of two heuristics (**H1** and **H2**) the following types of relations are defined:

- Equal (E): heuristics can be considered very similar or equal.
  Ex. The website heuristic H1: "Is the user provided with the essential information to carry out each task?" is the same as H2: "Only show essential information" for mobiles.

- More general (MG): H1 is more general than H2.
  Ex. The website heuristic H1: "Are the same elements grouped and located in the same place?" is more general than H2: "When designing an application, optimize edit view for data entry, grouping related items and prioritizing more commonly edited items at the top of the screen" for mobiles.

- More specific (MS): H1 is more specific than H2.
  Ex. The website heuristic H1: "Are the required values always marked using the same method?" is more specific than H2: "Required fields are marked".

- Conflict (C): There is a clear contradiction between both heuristics.
  Ex. The website heuristic H1: "At the top and the bottom of the page is there information about where the users are located and the last page visited?" is conflicted with H2: "Do not repeat the navigation on every page" for mobiles.

- Superseded (S): One heuristic presents a superseded of the other one.
  Ex. The heuristic H1 from website security: "Use Audio CAPTCHA to prevent spam" is a superseded version of the guideline H2 from website security: "Use graphic CAPTCHA to prevent spam."

In all the examples above, the set of heuristics should be cleaned-up before use. If the relation is E, the solution could be use either H1 or H2, without distinctions but one of them should be removed to avoid redundancy in the subset.

If the relation is a MG, the solution is to choose H2 and then to remove H1 from the subset of the selected heuristic. In this case, it is better to use the most specific one, therefore, the heuristic "*When designing an application, optimize edit view for data entry, grouping related items and prioritizing more commonly edited items at the top of the screen*" is chosen.

If the relation is MS, the solution is to choose H1. Otherwise, the selected heuristic will be "*Are the required values always marked using the same method?*"

All the examples above are relatively simple to detect and to treat. However, when a heuristic is superseded (S) by another or it is in conflict with other heuristic, further analysis is required. The goal of our approach is to help project managers to specify systematically arguments and decisions.

The first step to solve the conflict is to align the two heuristics that are in conflict. The way to get the design rationale is to construct a TEAM diagram from the individual TEAM diagrams of the heuristic definition.

The element question in the TEAM notation should be provided by project managers as it is a factor dependent of the context of the project. The options can be provided either but the recommendations in the heuristic description or manually provided by project managers. Finally, the weights associated to UX facets and each heuristic

is depicted in the diagram. The rest of the resolutions of the conflict are done by project managers with the help of the tool DREAM.

When a solution is found, a diagram containing the solution is recorded. So in the future it would be possible to retrieve the solutions found in previous projects. The next section illustrates the whole procedure in a particular case study.

3.3.4.3 Case study

This section shows the results of a real case study issued from the Ubiloop project from the Paul Sabatier University. Ubiloop is concerned with the development of solutions for improving the quality of the environment of a city using mobile and information technologies. The approach proposed by Ubiloop is to offer an incident reporting systems that allows citizens to report incidents in their neighborhood that might affect the quality of life, such a potholes, broken street lamps, graffiti, etc. The requirements for this application include the use of web technology. Moreover, the application should run in whatever platform/or devices citizens might have at their disposal, which might include smartphones.

In this working scenario, mobile technology is an essential ingredient because it allows users to make a report just after problems have been detected, when all the details about the incidents are still fresh in the users' mind.

Thus, 173 conflicting heuristic entries were collected from different sources: our information repository and literature about heuristics for mobiles and incident report forms.

As we shall see, heuristics for designing captchas are obviously conflicting. This example was deliberately chosen for focusing the discussion on the process of selecting and describing conflicts among the heuristics rather than improving the knowledge of conflicts of this particular element of the design.

We also have identified some UX facets that are important for this kind of application and for this case study including usability, as everyone should be able to use the application and perform a report in minimal time. Accessibility is an important facet enhanced by regulations. In addition, security (as a dependability facet) becomes an important facet as the kind of application we have in mind can suffer attacks from spambots that can shutdown web servers with massive spams and/or reduce trust in the information collected.

Based on these requirements, we have searched the heuristics for the three applications domains that concern our project: web applications (from our repository of information), mobile (from literature) and incident reporting (from literature). The selection process is described below.

3.3.4.3.1 Selection of heuristics

There is huge set of references in the repository of information that can provide suitable recommendation for dealing with the design or evaluation of application domains. A first search reveals as many as 177 entries for heuristics including: 82 heuristics for websites from our repository, 84 new heuristics for mobile phones and 11 new heuristics for incident reporting forms.

A first analysis of these 177 entries reveals several overlaps and conflicts among heuristics. For example, there are 16 clashes between heuristics for building web applications and heuristics for building incident reporting forms. As many as 138 conflicts concern heuristics for the development of web and mobile applications. Finally, 19 clashes are detected between heuristics for incident reporting forms and heuristics for mobile applications. These numbers are better presented in Figure 3.12.

Therefore, project managers should compare if it is possible to apply all these heuristics or, on the contrary, they detect conflicts and they have to choose one or another. Therefore, the next step was to clean the list of heuristics by removing duplicated entries.

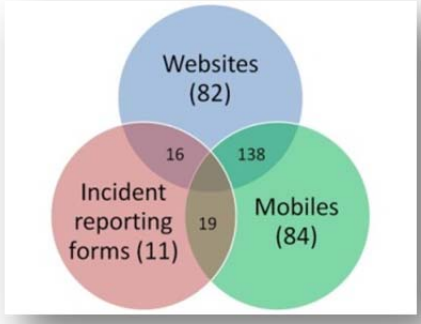


Figure 3.12 Overlapping of heuristics issues from three guidelines sets: web sites, incident reporting forms and mobile applications

3.3.4.3.2 Management of heuristics related to captchas

In our scenario, the project managers' goal is to develop an interactive system and for that they need a piece of advice which in our case it is provided by heuristics. The first question project managers have to answer is how users will have access the data on the website so that they will be able to complete the incident report form. The user should insert all the information of the incident and the system should assure the security of the transmission and the privacy of the personal information. In addition, the system should be able to control that only real users fill in the form. This means that the system should present the needed security restrictions to prevent spam or robot messages.

A possible solution to the questions above is to introduce in the user interface a new component to cover the requirement of the security: captcha. Captcha stands for "Completely Automated Public Turing test to tell Computers and Humans Apart"; it is a type of challenge-response test used in computing as an attempt to ensure that the response is generated by a person. Captchas help in ensuring that all reports have been inserted by citizens and not by bots because humans can read distorted text and or sound but current computer programs cannot. By asking users to fill in a form field the letter shown as distorted text/sound, it is possible to infer that the other fields in a form were duly completed by a human. Several possible implementations of captchas exist. One of the most striking conflicts can be translated as follows:

- **H1**: Prevent spams from bots.
- **H2**: Provide text alternative for non-textual elements.
- **Type of clash** (H1 in respect to H2): Conflict (C)
- **Rationale for describing the conflict**: providing a text alternative for non-visual captcha elements (as is the case of the *alt* attribute for images) will remove the security protection, but if alternative text is provided, programs or robots can also read the *alt* attribute from HTML pages.
- **Question**: How is the text presented in the graphic captcha?
- **Rationale for deciding the conflict**: importance of security versus accessibility and usability should help to decide if we keep (or not) captcha as a design option. Security is the most appropriate design option for implementing captcha (since this is necessary to describe which kind of accessible issues we are dealing with: visual impairments, audio, etc.).

The solutions for these conflicting heuristics requires a deep analysis of the associated trade-offs and the weight given to each factor can help project managers to make a decision. For example, since security is very important in our case study,

captchas should be implemented even if they can reduce the usability of the final user interface. If accessibility for blind people is important, then a visual-audio captcha should be used instead of a simple visual captcha. It is important to note that whatever the decision is, it will represent an infringement of some heuristics.

In order to ensure that the decision explicitly represents trade-offs, a design rationale is necessary. When conflicts among heuristics occur, designers resolve it and document the solution. To accomplish this, our approach proposes the use of design rationale. In Figure 3.13 how project managers can create a TEAM model for helping the decision-making process leading to the best design options for using captchas in the context of the Ubiloop project is represented.

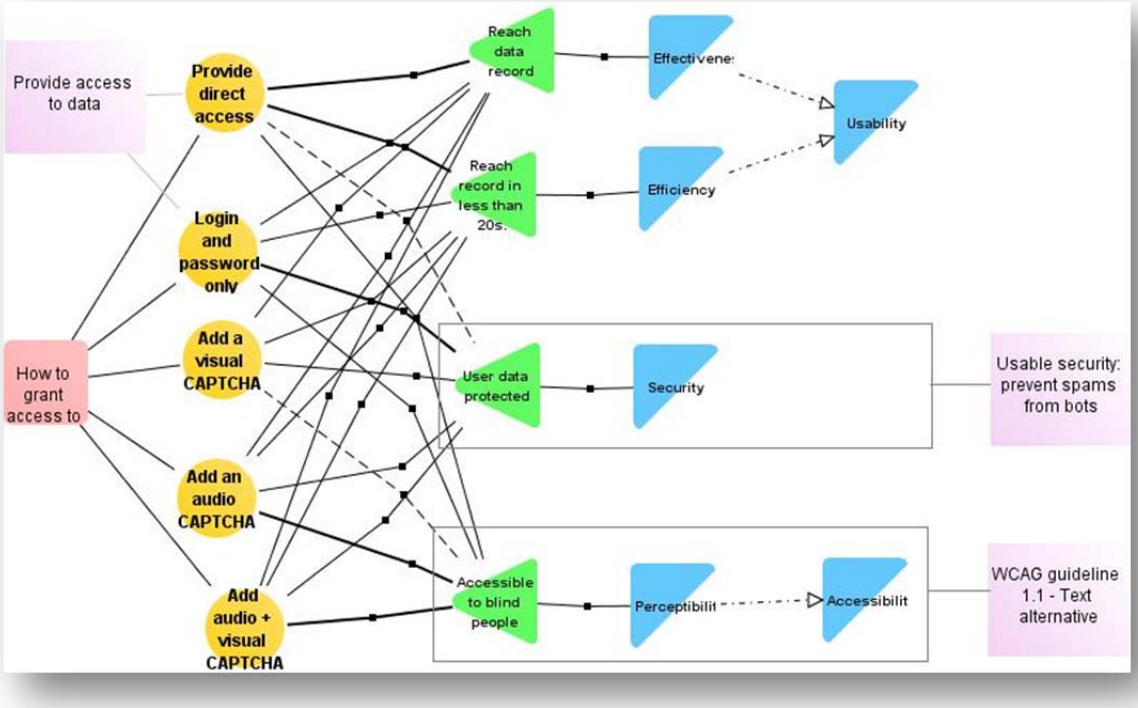


Figure 3.13 TEAM model describing the rationale for analyzing trade-off between captchas with respect to the requirements of the Ubiloop project and potentially conflicting heuristics

TEAM notation supports the observation of the relationships between heuristics and factors, thanks to its simplicity and readability, intended to be understandable by most of the actors involved during design (graphic designers, developers, customers, project managers,…). The weights (visible through the connecting lines) suggest that

the option *provide direct access* to the application does not comply with the security (dashed line in the connection) so this solution should not be selected. The many alternative implementations of captchas do favour security but only visual and audio-visual captchas are accessible for blind users, so that one on these options can be selected. This diagram explicitly shows the compromises that have been made between the heuristics for deciding the final solution to the problem.

Conflicting heuristics can be perceived at a glimpse at TEAM diagrams by looking at the lines connecting heuristics and design options. Our claim on "easier to observe" is based on the fact that a glimpse at a TEAM diagram allows us to detect the divergent 'favouring lines' when some paragraph of text would be required to explain the same thing. When diagrams are getting larger (in the case of multiple conflicts only) visualization techniques such as bi-focal browsers have been proposed [PAL07] as well as colour matrixes [MAR10].

3.3.4.4 Discussion

The main goal of the conflicting heuristics research is help users of large collections of heuristics (in particular designers and evaluators of interactive systems) to deal with trade-offs between conflicting heuristics. This work highlights these problems related to the management of conflicting heuristics. On the one hand we hope to deepen knowledge on the management of heuristics. On the other hand, we expect to help designers to understand the uses and misuses of heuristics. Indeed, heuristics are used throughout the world for providing guidance to projects but there is little evidence on how designers solve conflicting heuristics.

The approach presented is a possible solution but it imposes some constraints for the description of heuristics according to the TEAM notation. Nonetheless, it is not necessary to have all the heuristics systematically described to get the benefits. The description of heuristics can be done incrementally and the repository can contain entries that are not represented using the TEAM notation.

Moreover, we suggest that conflicting heuristics should be represented and documented only when they occurs in real projects for two main reasons: first of all because we need the contextual elements given by the real project to decide the best option; secondly, because the modeling activity can be time-consuming so it is better to make an effort when we can have an immediate benefit.

It is noteworthy that our approach for describing heuristics can be triggered either if we need heuristics for the design and/or evaluation of interactive systems.

However, in some cases, the solution of conflicting heuristics will be achieved after several iterations in the development process of the application.

For example, if we are in a design phase, we can select the heuristics, and we can provide a TEAM model to complete the description of individual heuristic and, possibly detect that heuristics might be in conflict with other ones. However, it might happen that the resolution of the conflict could not be carried out at the time of design as it might require some user testing to decide the trades-offs and the arguments allowing to solve the conflict. Thus, only when evaluations have been performed we can go back and record the solutions for the conflicts identified previously. The decision-making process leading to the resolution of the conflicting heuristics is done by project managers in an ad hoc manner. However, the description of the solution should be systematic and exhaustive.

In next Chapter 4, conflicting heuristics are implemented in Open-HEREDEUX and a real study case and its validation through a user test are detailed.

3.3.5 Algorithm of suggestion

Up to this point distinct research to get the most suitable set of heuristics has been detailed. The algorithm of suggestion joins the goals of the design or evaluation, the types of heuristics according to the receivers, the chance of choosing any financial constraints and the possible documentation of conflicting heuristics into an algorithm to get the most appropriate set of heuristics for a specific interactive system.

The suggestion of heuristics that will be used for designers and evaluators presents the same process with small different details for each user profile (designer or evaluator). Next, the six steps are explained (differences corresponding to each user profile will be explained when it will be necessary):

Step 1. Selection of the type of interactive system (to be evaluated or designed).

Step 2. Fix the goals to be achieved in the design or evaluation (functionalities, components, features, facets or ISO/IEC 25010:2011 attributes).

Step 3. Select the heuristic receiver (for designing or evaluating purposes).

Step 4. Refine the amount of the heuristics depending on the financial constraints.

Step 5. Manage conflicting heuristics.

Step 6. Obtain the set of heuristics as suitable as possible for our particular situation and use it.

The following details each step. Figure 3.14 shows a diagram of the process. Mandatory steps are marked using an asterisk (*):

Step 1*: The first step is the choice of the type of the interactive system.

Whoever will use the heuristics should determine the type of interactive system to be designed or to be evaluated; it will enable a more accurate set of the advised heuristics. At present, there are heuristics for three types of interactive systems: websites, virtual assistants and public kiosks.

Step 2*: The second step is the selection of the main goals.

Understanding that the main goal includes the specific aspects of the interactive system (functionalities, features and/or components), UX facets (dependability, accessibility, usability, playability, communicability, useful, desirable, findable, cross-cultural, plasticity and emotional) and the attributes of the standard ISO/IEC 25010:2011 that can be considered in a specific design or evaluation.

Therefore, at this step, the designer or the evaluator should know what the main goal to be considered is.

If the main goal is related to the specific aspects of the interactive system, the complete list of functions, features and components that correspond to this type of interactive system are presented. If the evaluator or designer chooses the UX facets as the main goal, the set of UX facets that the preselected type of interactive system has are given. Finally, if the option of attributes of the standard is selected, the list of 42 different attributes is offered.

Obviously, all this information can be listed because, previously, someone else had inserted and related this information according to the ontology detailed in section 3.2.

The designer or evaluator should read the specific aspects or facets or attributes of the standard and he/she should select or check the specific options that he/she would like to consider.

Step 3*: The third step is the selection of the receiver of the heuristics. The administrator chooses if the heuristics are for designing or evaluating purposes.

Step 4: This step is optional and it allows refining the final amount of heuristics according to: the UX degree and the time that could be invested in the evaluation.

Therefore, the level of UX that he/she would like to obtain is selected. Obviously, at least U degree heuristics will always be listed because it is the minimum amount of heuristics that are considered essential for getting a good design. This UX degree delimitation is shared by designers and evaluators.

The other limit that could be managed is the time spent (or needed) in each evaluation. Obviously, this limit is only applicable for evaluating purposes. Each evaluator can spend one minute to score two heuristics.

In summary, if the heuristics are for the designer, it is possible to delimit the amount of heuristics using the UX degree. After this delimitation, the process jumps to step 5. However, if the heuristics are for the evaluator, the heuristics can be delimited using the UX degree and the time restriction. In any case and considering that the minimum set of heuristics is the heuristics for the U degree, the minimum time is always half of the amount of U degree heuristics. Then, the process goes to the next step.

Step 5: Manage conflicting heuristics. At this point, the opportunity to check conflicts among heuristics appears. So, the possibility of choosing some related heuristics and deciding which one is more appropriate for the specific case is presented. This process enables the last refinement of the set of heuristics for a specific interactive.

Step 6*: With all previous information, the most adequate set of heuristics are listed. As we commented above, heuristics are written as sentences for the designers and as questions for the evaluators. The list is presented using different levels. The first level is the UX facet. The second is the UX degree and each UX degree presents its heuristic list.

Here it is possible to perform the last manual and individual selection of heuristics. If after the suggestion process there is some specific heuristics that do not fit with the goals, it is possible to uncheck them and these unchecked heuristics will not be saved to the final set.

Finally, you can use the set of suggested heuristics for a design step or for an evaluation. In the following Figure 3.14, a diagram of the whole process is presented. Furthermore, and bearing in mind that the suggestion algorithm is one of the most important contribution of the PhD, the following flowchart, Figure 3.15, details the internal process for getting the most appropriate heuristics for a specific interactive system.

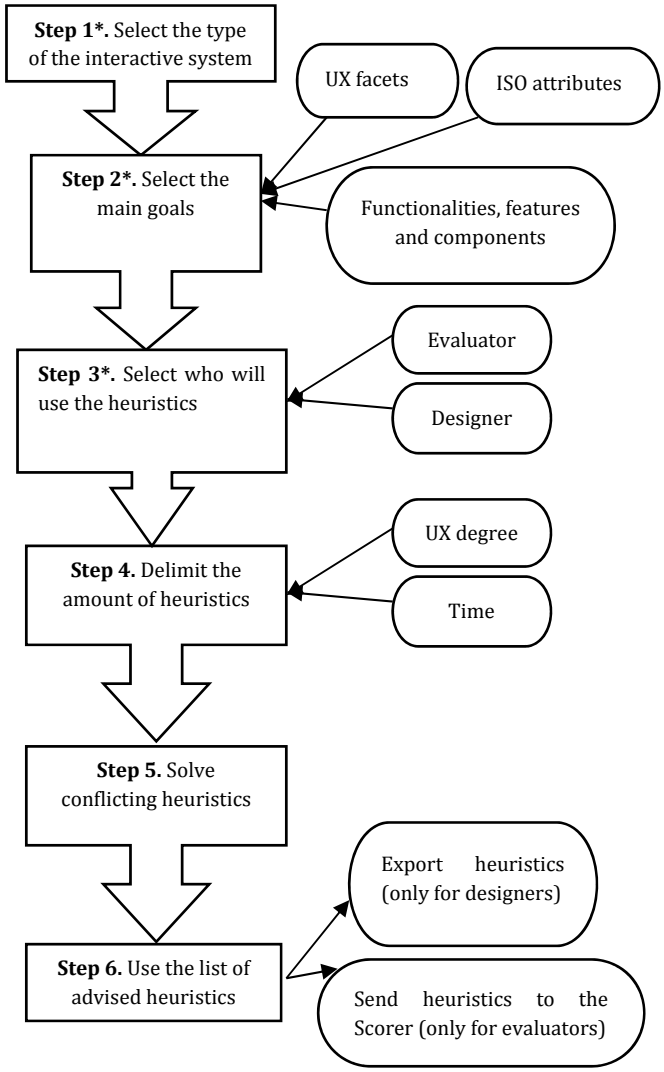


Figure 3.14 Diagram for a heuristics process suggestion

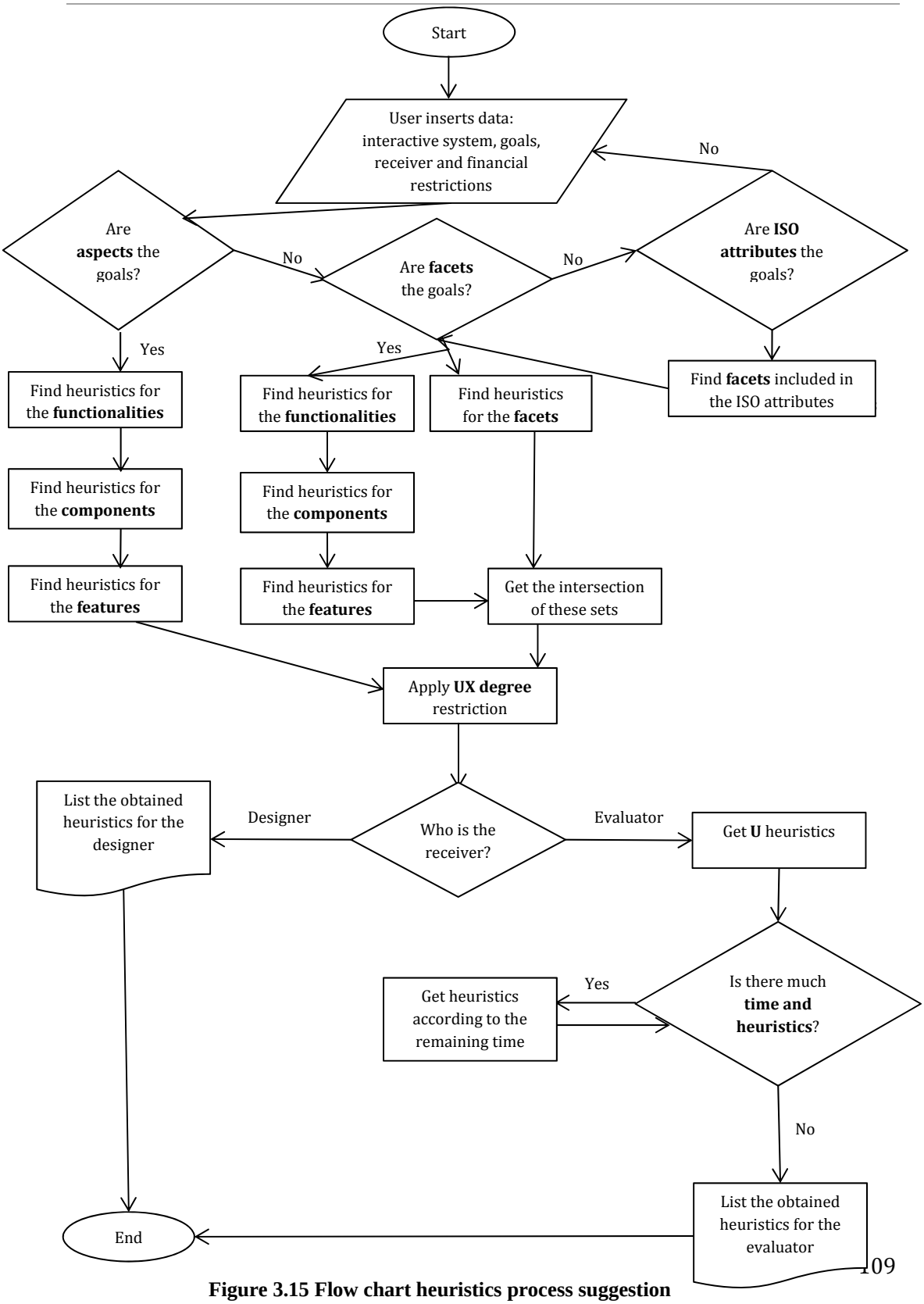


Figure 3.15 Flow chart heuristics process suggestion

3.4 Execution of the evaluation

Although the previous two sections present the solution of two different deficiencies in the heuristic evaluation methodology, this section describes how to support the realization of the evaluation per se. It is done following the heuristic evaluation methodology but including the realization of the evaluation in the methodology based on heuristics.

In section 3.2 the information needed for getting the most suitable heuristics for a specific interactive system is presented. At this point, it is necessary to present new information needed that is processed in the realization of the evaluation of different interactive systems. Furthermore, the severity factors selected for the process are also detailed. Finally, the simple process for scoring the set of heuristics for a specific evaluation is described.

3.4.1 Information needed

The information needed to carry out the realization of the evaluation is: projects, evaluators, severity factors and rating scale of these severity factors.

3.4.1.1 Project

A project is defined to carry out the evaluation of a specific type of interactive system. Therefore, in any case, the project is related to a type of interactive system and the list of heuristics suggested through the algorithm of suggestion.

3.4.1.2 Evaluators

The evaluator is a user role that includes two different kinds of evaluators: expert evaluators and end users.

Expert evaluators are those that know very well the heuristic evaluation methodology and how to run it. An expert evaluator knows the different phases of the methodology and the heuristics used in the evaluation. Usually, the expert evaluator is an HCI practitioner who is used to carrying out different techniques for considering the UX in the development process of an interactive system. Expert evaluators can also have knowledge about the interactive system but it is not a requirement to become an expert evaluator.

End users will use the system when it is ready. If UX would like to be considered in the development process of an interactive system, it is obvious to consider who will

use the system in the early stages of the development process. Therefore, an end user will evaluate the interactive system to get the emotional facet of UX definition. End users will score the most subjective heuristics to get their feeling, likes and dislikes about the interactive system.

3.4.1.3 Severity factors and their rating scale

Other important information in the realization of the evaluation is severity factors. The evaluators use the two most used severity factors to score the heuristics [NIE90].

Impact: The impact of the problem if it occurs: Will it be easy or difficult for the users to overcome? The rating scale used for the impact is:

- I do not agree that this is a problem at all.
- Cosmetic problem only: need to be fixed unless extra time is available on project.
- Minor problem: fixing this should be given low priority.
- Major problem: important to fix, so should be given high priority.
- Catastrophe: imperative to fix this before product can be released.

Frequency: The frequency with which the problem occurs: Is it common or rare? The rating scale used for scoring the frequency is:

- Never.
- Rarely.
- Sometimes.
- Usually.
- Always.

3.4.2 Scoring process

The scoring process is based on the basic process of the heuristic evaluation methodology. Each evaluator uses the projects to score the set of heuristics of each project. Evaluators score the set of previous suggested heuristics using the severity factors and their rating scale. In addition, they can write down some information to clarify the meaning of the score.

This step of the methodology does not present any aspect that is different from the original heuristic evaluation methodology. This is a manual process that is carried out in an easy way because the scoring process uses the information from the other

parts of the methodology. In addition, all the information inserted by evaluators is stored to reach results in the next stage of the methodology.

3.5 Results of the evaluation

As I detailed before, by default the heuristic evaluation provides qualitative and formative results. The proposal does not rule out this option, qualitative results should be provided in a UX evaluation. Moreover, I focused efforts in order to provide UX reports with a significant added value including quantitative and summative interpretation results. Combining qualitative/quantitative and formative/summative will enable the possibility to compare results of different evaluations and will provide detailed information to solve the problem faster.

Therefore, new measures are proposed and detailed in the following sections: correlation among users, the automatic classification of the problems, the UX degree and the percentage of UX.

However, before defining every measure, it is necessary to detail how evaluators should score each heuristic and previous basic statistical concepts.

Therefore, firstly, evaluators use the two most commonly used severity factors to score the heuristics. These are detailed in section 3.4.1.3. According to these severity factors and its rating scales, the following sections will show the proposed measures. It is important to highlight that the proposed measures work using the severity factors detailed in the previous paragraph, but it will also be possible to use them using others rating scales .

Secondly, correlation and hierarchical clustering analysis are used as the statistic base for the next UX measures.

In general statistical terms, the correlation among variables measures “how well the variables are related among themselves”. There are many measures to get the correlation but the most common one is Pearson’s Correlation [PEA95].

Pearson’s definition of correlation is: two organs (understanding organs as any measurable characteristic of an organism) in the same individual, or in a connected pair of individuals, are said to be correlated, when a series of the first organ of a definitive size (understanding size as the quantitative value of the organ) being selected, the mean of sizes of the corresponding second organ is found to be a function of the size of the selected first organ. If the mean is independent of this size,

the organs are said to be non-correlated. Therefore, Person's correlation is interpreted through three types or levels of correlation: high correlation [0.5...1] or [-0.5...-1], medium correlation [0.3...0.5] or [-0.3…-0.5] and low correlation [0.1…0.3] or [-0.1…-0.3].

Furthermore, the clustering methodology performs a hierarchical cluster analysis using a set of dissimilarities for the "n" objects (in our case, 4 objects) being clustered. At the beginning, each object is assigned to its own cluster. Then the algorithm iteratively proceeds joining the two most similar clusters until there is only one cluster. At each step distances between clusters are recalculated by the Lance–Williams dissimilarity [LAN67].

Specifically, due to the main aim of this type of clustering is the non-overlapping among clusters, Ward Hierarchical Clustering (WHC) [WAR63] is the particularly used clustering method. In our context, it means that one heuristic can only be part of one cluster. In addition, the WHC does not put together groups with smallest distance, but it joins groups that do not increase a given measure of heterogeneity too much. The aim of the WHC method is to unify groups so that the variation inside these groups is not increased too drastically. This results in groups in clusters that are as homogeneous as possible.

The Euclidean distance is the metric used to calculate the distance. The Euclidean distance is defined as the straight line distance between two points and it is used to decide the insertion of one heuristics in one or another cluster [DEZ09].

Let us see how to apply this statistical measure in the evaluation results.

3.5.1 Correlation among evaluators

The main objective of the Correlation among Evaluators (CaE) is to reach the reliability of the results. If evaluators score heuristics with similar scorings, the correlation value will be near '1' and it means that the results will be more reliable. On the contrary, if evaluators score heuristics in an opposite way (near -1) or without any kind of relation (near 0), the correlation will be wrong and it means that more evaluators are needed to get more reliable results.

Then, we define CaE as "the relation among the scorings that evaluators choose for a specific set of heuristics".

Taking the specific values of impact and frequency CaE is calculated. It is calculated grouping the evaluators in twos but in any case, CaE (according to the Pearson's correlation values [PEA95]) presents a value between -1 and 1.

Therefore, CaE is interpreted through three types or levels of correlation:

- High correlation [0.5…1] or [-0.5…-1]
- Medium correlation [0.3…0.5] or [-0.3…-0.5]
- Low correlation [0.1…0.3] or [-0.1…-0.3].

If CaE presents a high correlation, there are two possible interpretations.

- If it is [-0.5…-1] means that evaluators score the heuristics in an opposite way. Therefore, this is the worst result that an evaluation could present. It occurs when one evaluator scores the heuristic as OK and another one scores the heuristic as a catastrophe.
- If CaE is [0.5…1], this means that the evaluators score the heuristics with very similar scorings. This is the optimum value that the CaE should present.

Then, if CaE presents a low correlation [0.1…0.3] or [-0.1…-0.3] means that there is not a defined relation in the evaluation scorings.

Finally, if CaE presents a medium correlation [0.3…0.5] or [-0.3…-0.5] means that the relation is blurry. So, the results are not concluding.

This interpretation presents two main benefits in the results of a methodology based on heuristics:

- The first benefit is the confidence of the evaluation. If CaE is around 1, it means that all evaluators scored the heuristics in a very similar way. It means a strong UX agreement that increases the reliability of the results.
- The second benefit (and bearing in mind the previous one) lies with the budget of a real project. Usually it is difficult to convince people that the evaluation needs a larger budget because more evaluators are needed due to the type of interactive system, the kind of goals of the evaluation or the experience of the evaluators themselves. An important argument to convince the client to spend more money in this specific task of the project is that, with more evaluators, the project CaE would increase; assuring more reliable results. Consequently, the interactive system would acquire a higher quality level.

In summary, CaE is a quantitative and summative measurement to know the reliability of the results. In chapter 4 a real case study is detailed.

3.5.2 Automatic classification of problems

The main goal of the automatic classification of problems is to provide more understandable results.

> Four options are proposed in order to automatically separate the problems encountered during the evaluation. Up to now, the four proposed options are: **minor problem, medium problem, serious problem and catastrophic problem**.

The automatic classification is carried out by a clustering methodology. It is also important to highlight that to get the set of clusters the WHC statistic methodology is used.

This automatic classification provides the results of the evaluation with two important advantages. The first advantage, and the most important, is the elimination of the post-evaluation meeting. Usually, and according to our experience, the post-evaluation meeting is exhausting, mainly when the set of heuristics is large or the evaluators do not have much experience.

In addition, this meeting must be held not too long after evaluations, so evaluators need to remember why they scored each heuristic with a specific score. Using WHC the heuristics are classified automatically in different gravity levels, thus avoiding many hours of discussions about the best score for specific heuristic.

The second advantage is in reference to the interpretation of results. When the client receives the results, it is very important to highlight the most important problems as well as the best aspects of the interactive system. Therefore, the results should be presented in as clear and simply way as possible. Due to the amount of different scorings that can appear for each heuristic, it is very difficult to explain to the client differences among some ratings.

For example, if a heuristic is scored as “Major problem (Impact) that always happens (Frequency)”, how can the client note whether this first heuristic presents a worse problem than another heuristic scored as “Catastrophe (Impact) that usually happens (Frequency)”. The specific feature for deciding which of these two cases should be solved first is not obvious. The consideration of the WHC solves this problems dividing the whole set of scorings into 4 gravity levels (in these 4 levels the heuristics that are not a problem are not considered).

To sum up, the automatic classification of the problems is a qualitative and summative measure to enable a better comprehension of the results.

3.5.3 Percentage of UX (PUX)

The principal goal of the Percentage of User eXperience (PUX) is to quantify the amount of UX that an interactive system has.

```
Then, PUX represents the amount of heuristics which are
not a problem regarding the total of heuristics.
```

PUX is presented for the heuristics of each UX degree and for the total amount of heuristics. Specifically:

$$PUXi = 100 - \left(\frac{\sum HPi * 100}{\sum HTi}\right)$$

Where "i" represents the UX degree (U, UU, UUU). HPi is the amount of heuristics of the "i" degree (U, UU, UUU) that presents some kind of problem. HTi is the total amount of heuristics of the "i" degree (U, UU, UUU).

Then to get the total percentage of UX problems the PUXt formula is presented. The main point of the PUXt formula is that all UX degrees have the same importance in the total percentage of UX problems. In other words, the degree with more heuristics is not more important than another degree with less heuristics. The number of heuristics that each UX degree presents is not considered for this formula. The three UX degrees have the same importance for reaching the percentage of UX problems. The PUXt formula is:

$$PUXt = 100 - \left(\frac{\sum Pu * \frac{100}{3}}{\sum HTu} + \frac{\sum Puu * \frac{100}{3}}{\sum HTuu} + \frac{\sum Puuu * \frac{100}{3}}{\sum HTuuu}\right)$$

Where Pu is the number of problems of the U degree and HTu is the total number of heuristics of the U degree. PUU is the number of problems of UU degree and HTuu is the total number of heuristics of the UU degree. Finally, the Puuu is the number of

problems of UUU degree and HTuuu the total number of heuristics of the UUU degree.

Briefly, this measure is a quantitative and summative measure to get the level of quality of specific interactive system.

3.6 Conclusions

A **repository of information is essential for reaching an optimum process for suggesting heuristics**. Therefore, heuristics are the most valuable data that should be related to the components, features, functionalities and UX facets. Using this relation and considering that an interactive system is defined through components, functionalities, features and UX facets, the selection of the most suitable heuristics for a system is evident.

Regarding the **suggestion for heuristics**, it **solves one of the main deficiencies of the heuristic evaluation methodology**. In addition, the methodology detailed in this section for the suggestion of the best heuristics for a specific interactive system includes different considerations such as different goals for design or evaluation, different types of heuristics according to the receiver, financial constraints and the possibility of documenting conflicting heuristics. All **these parameters enable the inclusion of this methodology in many different enterprise contexts** where many different kinds of interactive systems are developed.

Then, these two first stages of the methodology provide the needed conditions to facilitate the evaluation process per se. This stage is not really automatic or optimized but due to the consideration of the rest of the methodology, **the realization of the evaluation per se become a very easy step of the methodology based on heuristics**.

Finally, the measures for acquiring UX evaluation results are presented in different ways and according to different goals. But in any case, the most general goal is to **facilitate the interpretation of these results** to improve the specific interactive system and to **compare these results in the possible following evaluations of improved versions and other systems, for instance systems from rivals**.

Therefore, as is detailed in previous sections, the main goals of the UX measures are: **to get more reliable results** through the best number of evaluators, **more understandable results** translating the wide range of results in four gravity levels,

the **standardization of the results** for comparing the results of different versions of the same product or the same kind of interactive systems, and the quantification of the amount of UX reached in a specific interactive system to know if the interactive system resulted in a positive experience for end users.

Obviously, the **typical qualitative results are not ruled out**. This set of UX measures should be presented including the usual qualitative information such as the needed improvements to be applied in the interactive system.

I recommend the presentation of the improvements using the UX degree measure. Due to its nature of dividing heuristics in different levels of importance, the presentation of the improvements through the UX degree permits a hierarchical method to present the most important and urgent problems to be solved first.

In addition, the presentation of the results is proposed through a standard format. It means, as you shall see in the following chapter, that **the results report for the methodology based on heuristics is presented using the ISO/IEC 25062:2006 standard** (*Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Common Industry Format (CIF) for usability test reports*).

Chapter 4

"Seeing is believing"
"Cal veure-ho per creure-ho"

Open-HEREDEUX: application of the methodology

4.1 General description

Based on the methodology described in Chapter 3, OPEN-HEREDEUX is presented as a solution to consider the UX in an interactive system design or evaluation process. OPEN-HEREDEUX is the short name for our proposal: "OPEN HEuristic Resource for Designing and Evaluating User eXperience".

It is a resource that will enable UX experts to design for UX and evaluate UX in a semi-automatic way. Therefore, following the four main aspects of the methodology, Open-HEREDEUX includes the next four components:

Open Repository (Figure 4.1- **1**) is provided with all the necessary information to achieve the set of heuristics as complete and minimum as possible. Therefore, it presents the repository of information part of the methodology. In section 4.2, Open Repository is detailed. Moreover, how to set new information is also presented.

Adviser of heuristics is the second component (Figure 4.1- **2**). It intends to be a tool whose objective is to propose, for a specific interactive system, the most appropriate list of heuristics to be used. It is suitable for such usages as recommendation principles in a design phase or as evaluation principles in a UX evaluation based on heuristics. Therefore, it includes the heuristic suggestion part of the methodology described in section 3.3. Section 4.3.1 reports two specific cases where Adviser lists different heuristics and how conflicting heuristics are documented.

Therefore, heuristic suggestions can be used either as a list of recommendations to design a specific interactive system or as an input for the next, and third, component: **Scorer of heuristics**, which is in charge of carrying out the realization of the evaluation (Figure 4.1- 3). This is explained in section 4.4.

Finally, **Results Analyzer** is the last component and it is presented in section 4.5. It provides quantitative/qualitative and formative/summative data interpretation (Figure 4.1- 4).

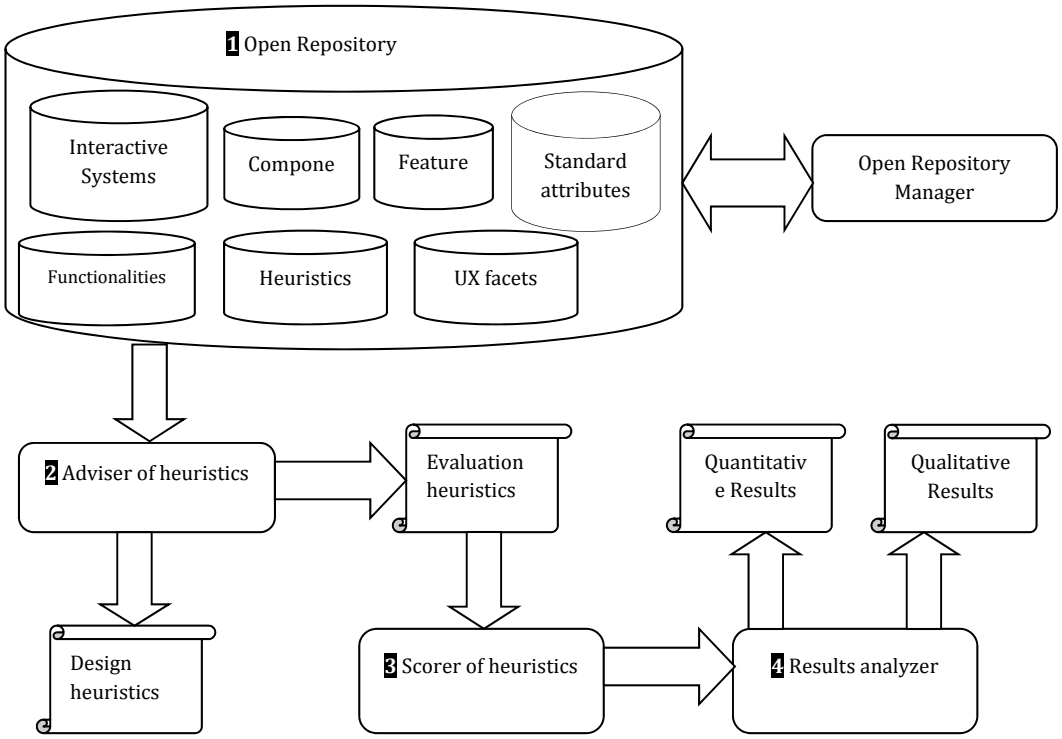


Figure 4.1 Open-HEREDEUX overview

In summary, Open Repository collects all the needed information that Adviser uses to list the set of heuristics as appropriate as possible for designers or for evaluators. Then, if an evaluation should be carried out, Adviser sends the list of the most appropriate heuristics to Scorer. It will serve the evaluators for scoring all the heuristics. This module also saves all the information to be sent to Results Analyzer. Finally, Results Analyzer is the part of the system where results are processed.

Figure 4.1 shows the complete Open-HEREDEUX overview.

Open-HEREDEUX is available on this url: www.grihotools.udl.cat/openheredeux. Figure 4.2 shows the main page. In following sections, details about every component are described.

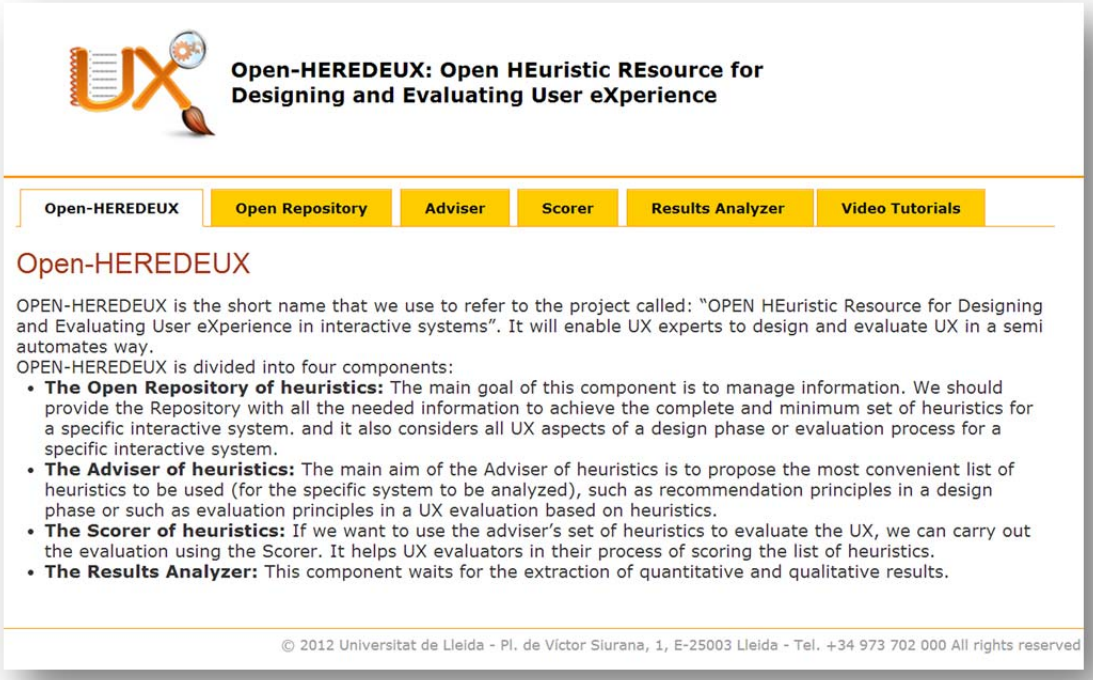


Figure 4.2 Open-HEREDEUX main page

4.2 Open Repository and its manager

Open Repository (Figure 4.1- 1) and its manager are the implementation of the repository of information part of the methodology. The manager shows the essential interface to manage all the needed information to use Open-HEREDEUX.

Open Repository Manager enables **two main tasks** using the information of Open Repository: **to look up information and to manage heuristics relations**. Specifically, it allows in an open way:

- Look up heuristics for each functionality, feature, component, attribute of the standard and UX facet.

In addition, if one has administrator rights, it is also possible to:

- Insert and update heuristics, functionalities, features, components, attributes of the standard, UX facets and, in addition, the information for the interactive systems.
- Relate heuristics to different functionalities, features, components, attributes of the standard and UX facets.
- Relate functionalities, features, components, attributes of the standard and UX facets to interactive systems.
- Insert and update the options of severity factors to be used in the evaluation process.
- Manage projects.
- Manage user permissions.

Open Repository should be provided with all the needed information to achieve (in Adviser) a set of heuristics that is as complete and minimum as possible for a specific interactive system. It should also consider all aspects of UX either in its design phase or evaluation process for a specific interactive system.

The most important advantage is that it presents a wide pantry (or repository) of heuristics. Therefore, it provides a single source of heuristics where it is possible to search and find the most adequate set of heuristics for different interactive systems.

The second benefit is that Open Repository can be used by designers and evaluators. In addition, heuristics are useful in any step of the development process. An open characteristic is another important advantage. It implies the constant evolution of Repository and open access for everybody.

Furthermore, the relations among heuristics and functionalities, features, components, attributes of the standard and UX facets enable the list of the best set of heuristics from Adviser.

And finally, although Open Repository takes part in Open-HEREDEUX, it can work as a standalone, regardless of the other components. Figure 4.3 shows the public functionality of Open Repository.

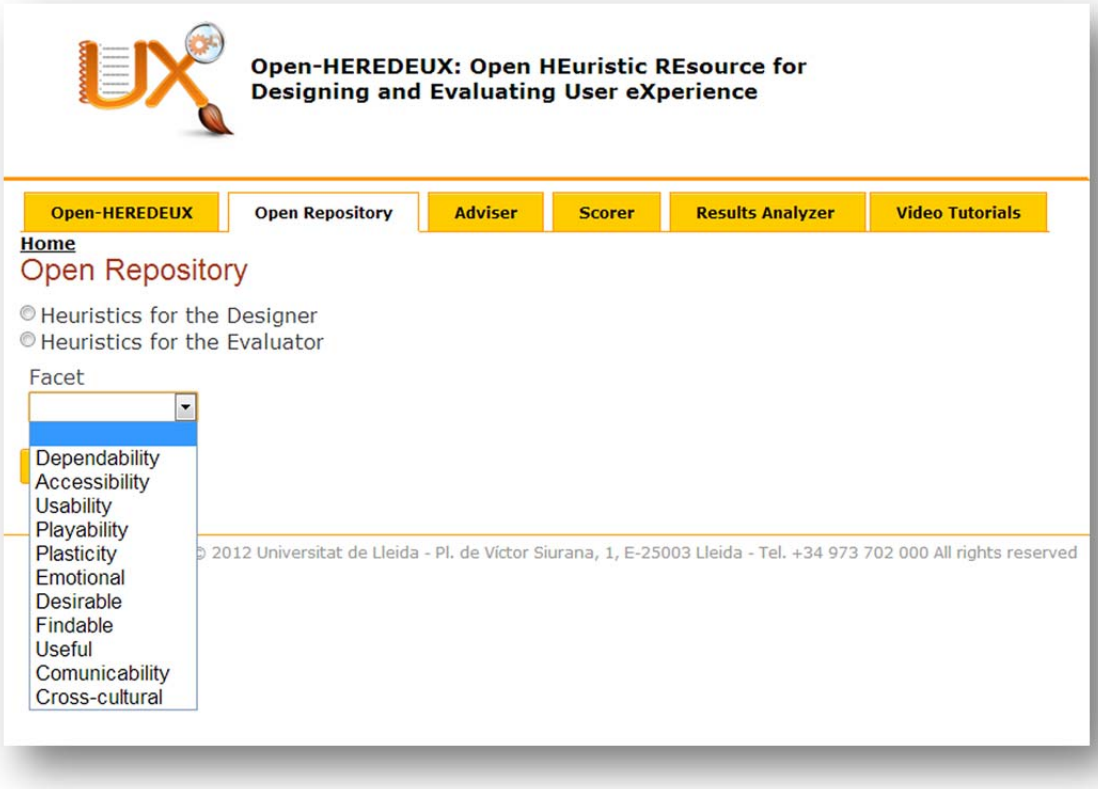


Figure 4.3 Open Repository main page

4.2.1 Setting new information in Open Repository. iTV case study

This part of the project is carried out together with Andres Solano, a PhD researcher from Universidad del Cauca (Colombia) who developed his Master's Degree thesis on the topic: "*Methodological proposal for the collaborative evaluation of the usability of interactive digital television applications*". He is defining a new methodology for

evaluating the usability of interactive television applications in a collaborative way. Bearing in mind that one of the methodologies that he would like to improve is the heuristic evaluation and the strength relation between his advisors (César A. Collazos and José Luis Arciniegas) with our research group, the proposal of setting his information/heuristics in Open Repository was happily and quickly accepted by everybody.

Interactive Television (iTV) is the convergence of television and computer technologies. Usability evaluation for applications based on emerging information technology brings new challenges. The main iTV feature is that the user may interact with the application; therefore usability should be a main concern when designing iTV applications. Current research usually focuses on iTV applications from a technical point of view, rather than a user-centered approach. There is a need for new usability evaluation methods or at least for the use of traditional evaluations in novel ways [SOL12].

iTV includes relevant aspects such as: ease-of-use, entertainment, information, among others. Ease of use is clearly a priority for design interactive applications. Many users are accustomed to using the TV in a passive way (channel changes only), so, the new paradigm consists in designing interactive applications that are more intuitive and clear as possible, trying to get users to become an active part of the TV schedule [SOL12].

The process involved for introducing new heuristics in Open Repository is scheduled in 5 main tasks:

1. **Discard duplicate heuristics.** Initially, in [SOL12], a set of 164 heuristics for iTV applications had been proposed. However, before including some new heuristics to Open Repository, the heuristics already included in it should be reviewed manually to avoid duplicates. Obviously, some iTV heuristics are the same as those referred to with other interactive software systems, and then it was necessary to manually decide which heuristics for iTV were already registered in Open Repository, and not include this repeated heuristics in Repository again.
   However, it is important to remember that the heuristics that are already included in Repository should be related to this new and previously created interactive system. This task was more difficult than originally planned because repeated heuristics should be identified and also those with a high degree of similarity. Then we should decide which of them are discarded or included (with modifications).

In addition, the large amount of heuristics stored into Open Repository increases the difficulty of this task. Up to this moment, 363 heuristics were stored and comparing the new 164 iTV heuristics with all of them takes much time and it is error prone. At this point, 22 heuristics were detected as already included in Open Repository.

2. **Writing heuristics as declarative and interrogative sentences.** Some iTV heuristics were not written in a declarative or interrogative mode. The task of writing the heuristics in declarative mode (for designers) and interrogative mode (for evaluators) was carried out manually.

3. **Relationship of software and hardware components.** Table 4.1 presents the software components identified for an iTV application. New inserted components are marked with an asterisk (*).
   Based on Table 3.13 in Chapter 3, it is noteworthy that in the general group "Type of content" this component was added: *buttons*. Additionally, the general group was added: *Settings*. These components were added considering the physical features of the interaction in iTV environments.

General group	Specific feature
Type of content	Forms, tables, lists, dates, times, numerical values, money signs, buttons*.
Information	Pictures, news, graphics, format, text, URL, abbreviations, audio, nomenclature, colours, icons.
Data management	Information transmission, sign in form, log in form, updating information, information validation, information recovery.
Search	Search form, results of search.
Navigation area	Titles, cursor, shortcuts.
Error management	Error message
Emergency exits	Exits
Settings*	Restricted colors*, restricted sounds*.
Help and documentation	Help

Table 4.1 Classification of software components for iTV applications

4. **Relationship of UX facets.** The iTV heuristics are previously related to the facets: usability, accessibility and findable. Then, these heuristics were only related to these existing UX facets. This means that if some other facet would be considered in the design or evaluation process, a new insertion of information will be required due to the lack of information about other facets.

5. **Evaluation of the degree of relevance of the heuristics.** To define the UX degree of the iTV heuristics we made a consensus among seven experts with the following profiles: at least 3 years of experience in heuristic evaluations, knowledge about the areas of HCI and User-Centered Design and basic knowledge about iTV applications.

Each expert decided individually what UX degree best fits each heuristic. Then, a consensus on the UX degree of each heuristic was obtained regarding the selection of all experts.

It is also important to highlight that the information presented previously (components and facets) is related to the heuristics in order to determine the most appropriate set of heuristics for iTV applications. It is noteworthy that the process for determining which heuristics should be included in each component and facet was a manual process. In addition, to ensure the consistency of the heuristics, all of these are related to a single facet and one (or several) components.

Table 4.2 presents the number of heuristics for each facet entered in Open Repository. Before this insertion, Open Repository stored 363 heuristics divided into 7 UX facets. This iTV experience was provided with 142 new heuristics which were entered into Open Repository. They will contribute to the design and evaluation of iTV applications. Thus, nowadays, Open Repository presented 505 heuristics.

Facet	Number of heuristics		
	Before	iTV heuristics	Later
Reliable	27	0	27
Usability	239	123	362
Accessibility	15	4	19
Emotional	6	0	6
Findable	19	15	34
Communicability	41	0	41
Cross-cultural	16	0	16
Total	363	142	**505**

Table 4.2 Heuristics per UX facet

4.2.1.1 Discussion

As we have seen, setting new information in Open Repository is a handmade task, implying a huge effort to add foreign heuristics (of other interactive systems).

The most cumbersome tasks are to discard duplicate heuristics and to set the UX degree for these new heuristics. The process of rejecting duplicated heuristics needs the comparison between the set of new heuristics and the collection already stored in Repository.

Moreover, if the UX degree for the new heuristics would be evaluated to insert validated data, some UX or interactive system expert should be found to decide the best UX degree (U, UU, UUU) for each new heuristic.

Up to this moment, although the use of Open Repository is very useful for designers and evaluators due to the fast access and the amount of heuristics, the process of inserting new information is not an easy task (as we have seen in this iTV case). This difficulty comes from the need to relate the new heuristics with functionalities, features, components, attributes, facets and so on.

4.3 Adviser of heuristics

Adviser of heuristics (Figure 4.1-2) runs the algorithm for suggesting (to the designer or evaluator) heuristics. Adviser of heuristics offers the set of heuristics as adequate as possible for a specific interactive system. Therefore, according to the information stored in Open Repository, Adviser is able to list heuristics for two different receivers.

Firstly, user interface designers will be interested in the resource to get the most appropriate recommendations when they are designing a specific system.

Secondly, the evaluator who would like to analyse the UX of an interactive system will use Adviser to get the set of heuristics that should be considered in the evaluation of the interactive system.

Therefore, according to the suggestion algorithm, Adviser enables project managers the following tasks:

- Check the type of interactive system that you would like to get heuristics for.
- Fix the goals of the suggestion: choosing different aspects of the interactive system (components, functionalities and features), UX facets or attributes of the ISO/IEC 25010:2011.
- Select the receiver of the set of heuristics: the designer or the evaluator.
- Delimit the number of heuristics through the UX degree and time restriction (the last one only if heuristics are for the evaluator).
- Check and document conflicting heuristics.
- Download the heuristics for the designers or save the heuristics for the evaluator to use them in Scorer of heuristics.

Figure 4.4 shows a summary of the process in Open-HEREDEUX interface.

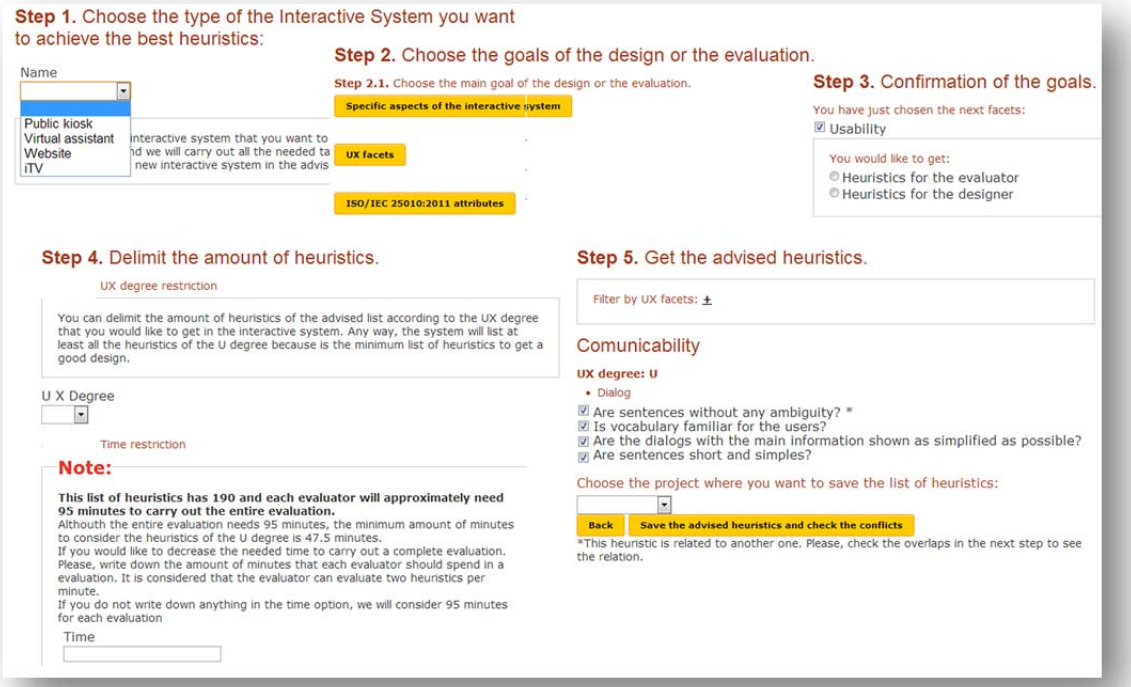


Figure 4.4 Adviser's procedure

This component has an important advantage when choosing the best heuristics for a specific interactive system: turning a manual process and a process based on UX expert experience into a semi-automatic process where UX experts will only review the set of heuristics suggested by Adviser.

Another benefit is the double usage of the heuristics. The whole set of heuristics that are available in Adviser can be used for designers and for evaluators. This means that Adviser can be included in two phases of the development process of an interactive system: in the designing and evaluation phases.

4.3.1 Example of using Adviser

Bearing in mind the real case studies presented in section 1.1, I would like to apply ISO/IEC 25010:2011 attributes in the design or evaluation (in this case it does not matter who will use the heuristics, the set of heuristics will be the same in both

cases) of these two interactive systems (virtual assistant and public kiosk) using Adviser of heuristics.

Adviser presents the set of attributes to allow their selection to be considered in a specific system. According to the choice of the attributes, Adviser (in a transparent process for end users) looks for the functionalities, features and components of the specific system. Then, it checks the heuristics that the functionalities, features and components have. This part is used to select the heuristics of the specific system that are included in each attribute of the standard and to reject the heuristics that are included in the attribute but are not for the specific system.

Therefore, although both systems (remember, virtual assistant –Berta– and public kiosk –CIP–) have the same main goal (to apply the standard ISO/IEC 125010:2011), the list of the most suitable set of heuristics is different. Table 4.3 presents the number of heuristics (of each facet) that each attribute has for each interactive system. As you can see in Table 4.3, the number of heuristics for the virtual assistant (Berta) and the public kiosk (CIP) is different. For some attributes, the virtual assistant has more heuristics and other times, the public kiosk has more. Although the whole set of ISO attributes presents heuristics from websites, public kiosks or virtual assistants, there are some attributes without heuristics due to the available heuristics for both interface does not cover all attributes of the standard.

ISO/IEC 25010 : 2011			UX facets	Number of heuristics	
				Berta	CIP
4.1 Quality in use	Effectiveness		Usability	92	143
			Playability	0	0
			Useful	2	2
	Efficiency		Usability	5	5
			Playability		0
	Satisfaction	Usefulness	Useful	2	2
		Trust	Emotional	5	0
			Playability	0	0
			Desirable	0	0
		Pleasure	Emotional	1	1
			Playability	0	0
			Desirable	0	0
		Comfort	Emotional	2	2
			Playability	0	0
			Desirable	0	0
	Freedom from risk	Economic risk	Dependability	5	5

ISO/IEC 25010 : 2011			UX facets	Number of heuristics	
				Berta	CIP
4.2 Product quality		mitigation			
		Health and safety risk mitigation	Dependability	2	2
		Environmental risk mitigation	Dependability	0	0
	Context coverage	Context completeness	Usability	107	139
		Flexibility	Usability	36	36
			Accessibility	1	6
	Functional suitability	Functional completeness	Useful	2	2
		Functional correctness	Accessibility	2	2
			Playability	0	0
		Functional appropriateness	Accessibility	3	1
			Playability	0	0
	Performance efficiency	Time behavior	Usability	5	5
		Resource utilization	Dependability	1	0
			Accessibility	1	1
		Capacity	Dependability	0	0
	Compatibility	Co-existence	Plasticity	1	0
		Interoperability	Accessibility	0	1
			Plasticity	0	0
	Usability	Appropriateness recognisability	Usability	127	157
			Findable	14	11
		Learnability	Usability	140	177
			Playability	0	0
		Operability	Usability	11	16
		User error protection	Usability	0	15
			Playability	0	0
		User interface aesthetics	Usability	128	176
			Playability	0	0
		Accessibility	Accessibility	2	4
	Reliability	Maturity	Dependability	17	12
		Availability	Dependability	2	3
			Accessibility	3	10
		Fault tolerance	Dependability	8	0
		Recoverability	Dependability	4	0
	Security	Confidentiality	Dependability	7	3
		Integrity	Dependability	9	3

ISO/IEC 25010 : 2011		UX facets	Number of heuristics	
			Berta	CIP
	Non-repudiation	Dependability	9	0
	Accountability	Dependability	0	0
	Authenticity	Dependability	3	0
Maintainability	**Modularity**	Dependability	2	2
	Reusability	Dependability	2	2
	Modifiability	Dependability	9	7
		Accessibility	1	1
Portability	**Adaptability**	Accessibility	1	4
		Plasticity	0	0
	Installability	Plasticity	1	0
	Replaceability	Plasticity	1	0

Table 4.3 Number of heuristics for each attribute in the specific case studies

4.3.2 Example of conflict management

As it is presented in Chapter 3, the conflict management is proposed for solving possible overlaps among heuristics; overlaps that can appear when users use more than one heuristic source. In this section how it is implemented in Open-HEREDEUX and its validation through a user test are presented.

Despite the process of documenting conflicting heuristics being optional and can be run on its own, Open repository stores the heuristics and the information obtained in the documentation phase. Then, Adviser implements the interface to check the possible conflicts and how to document each of them.

4.3.2.1 Comparative user test for the validation of conflict management

The main goal of this test is to evaluate whether the process for managing conflicts among multiple heuristic sources are valued positively by participants. The wide acceptance of this process by end users (considering end users as interface designers and evaluators or UX practitioners) will present a positive achieved challenge of this part of the project.

The comparative user test is carried out using two different but very similar versions of Open-HEREDEUX. In version one, version-1, of Open-HEREDEUX, the process for documenting conflicting heuristics is not implemented and in the other version, version-2, there is the process for documenting conflicting heuristics. Then, the hypotheses proposed for both versions are:

- H1. If multiple sources of heuristics are used, it is very likely that users will find conflicts among heuristics over time.
  - H1a. If people use multiple sources of heuristics in their work, they might have found conflicts among them in the past.
  - H1b. If people are using tools (such as Open-HEREDEUX that include multiple sources of heuristics), they might be able to find conflicts among heuristics.

- H2. If people are using tools (such as Open-HEREDEUX that include multiple sources of heuristics), they will be more efficient in recognizing, understanding and reporting conflicts that are clearly exhibited by the tool.
  - H2a. If there are conflicts among multiple heuristics sources, participants will find conflicts easier if they are documented at Open-HEREDEUX.
  - H2b. If conflicts among heuristics are represented using rationale notation, people will understand the conflicts without additional training on the notation.

- H3. If people are using tools (such as Open-HEREDEUX that include multiple sources of heuristics), they might be willing to report their solution or they might understand and trust the solution proposed by others users.
  - H3a. If people found a conflict among heuristics; they are willing to document their solution for conflict resolution.
  - H3b. If a tool (such as Open-HEREDEUX) shows solutions for conflicts among heuristics, people would trust those solutions.

Let us see how the test was carried out.

4.3.2.1.1 Participant profile

The test was carried out with 9 males and 3 females. The age of the participants is detailed in Table 4.4.

Age	20-30	31-40	41-50
Number of participants	3	5	4

Table 4.4 Age of participants

In addition, all of them had as their primary background Computer Science but one of them has also graphical design. Regarding their highest level of studies, there were 6 Doctors, 5 participants have a Master's Degree and one of them has a Grade.

Currently, 7 participants are professors in computer science and they were also professors in a master specialized in HCI, 3 participants are web developers and

software architects, 1 participant is a UX, usability and accessibility designer and HCI consultant. Finally, 1 participant is a graphical designer, web developer, information architect, UX, usability and accessibility designer.

In addition, they have experience in designing/evaluation user interfaces. 6 participants have 1-4 years of experience. 4 participants have 5-10 years of experience and 2 participants have more than 10 years of experience in this area.

4.3.2.1.2 Scenario, tasks and procedure

The scenario for both versions was:

"Let us assume that you are a designer of a company where many different types of websites are developed. One of your clients asks you to design a feature to control the access to some parts of their website using a log in. Your main task is to design such as log in. Importantly, one of the main client's requests is that your design solution should respect existing recommendations for ***usability****,* ***accessibility*** *and* ***security*** *of the websites. In order to make sure that your design option respects all those properties, you are invited to use the tool Open-HEREDEUX for getting the appropriate set of guidelines that might help you to accomplish your main task. "*

Then, bearing in mind both versions of the test, the tasks of the test were a little bit different. The common tasks for each version were:

- **Task 1:** Training task: Search heuristics for the usability evaluation of website applications.
- **Task 2:** Search heuristics for the evaluation of Log In functionality of the website application.
- **Task 3:** Save this set of heuristics in the project called "Conflicts test".

The additional tasks for version-2 were:

- **Task 4:** Check the conflicts among heuristics. Look at the information previously saved. And save it if you like this solution or you prefer another one.
- **Task 5:** Select two heuristics to insert a new conflict.

The procedure was the usual for user testing:

- Fill in user profile form
- Fill in previous experience questionnaire
- Carry out the tasks
- Fill in post-tasks form
- Fill in post-test form

4.3.2.1.3 Results

Results of the comparative user test are related with the efficiency of the process for documenting conflicting heuristics and the verification of the hypothesis presented above.

In Figure 4.5 and Figure 4.6, the time that each user spent in each task is provided for the test applied to version-1. As is shown, the longest time is no more than 5 minutes. Therefore, the process to prepare an evaluation that includes the selection of the most suitable heuristics for the specific case is very efficient.

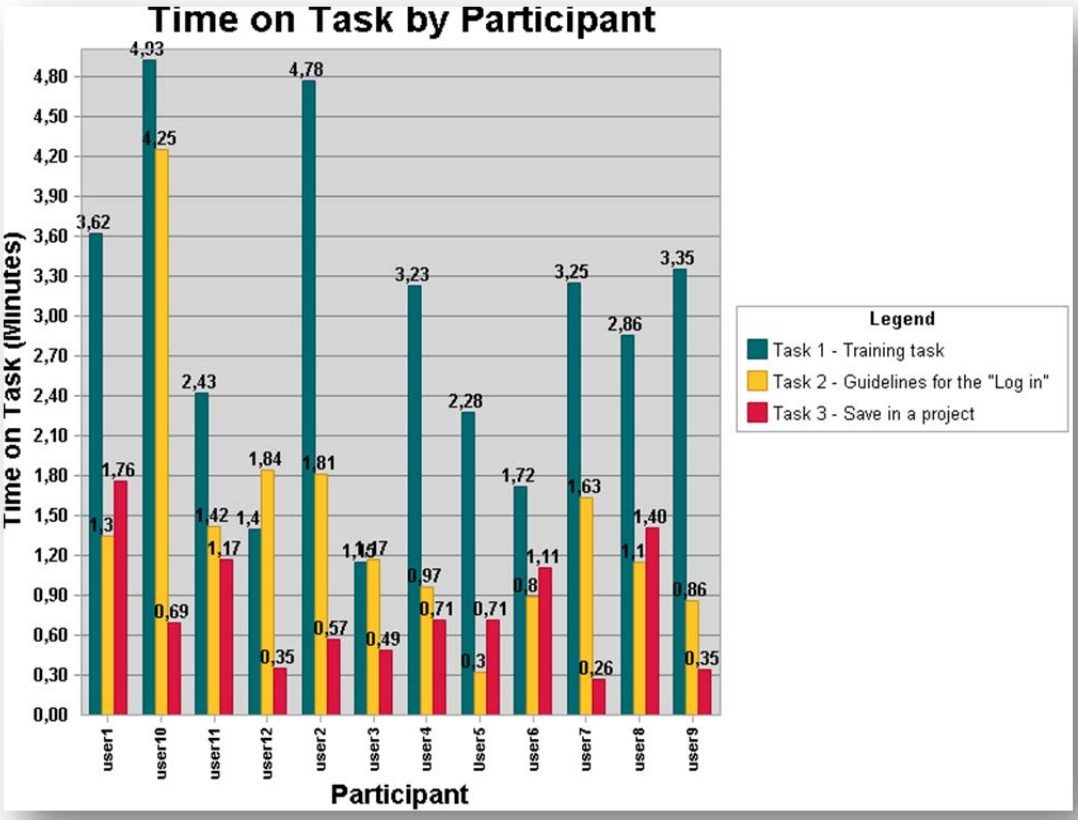


Figure 4.5 Time on task by participant (to version-1)

Furthermore, the time that each user needs to use version-2, that has the conflict management implemented, is presented in Figure 4.7**¡Error! No se encuentra el origen de la referencia.** and Figure 4.8. It is important to note that in this test the

most difficult task is task 4 because it is the task that presents the interface for the conflict management. Therefore, the first time (and as is detailed in the following sections) that users use this interface they have to know it and understand the information. However, in the next task (Task 5) the same tasks take much less time due to users already know the interface and how it works.

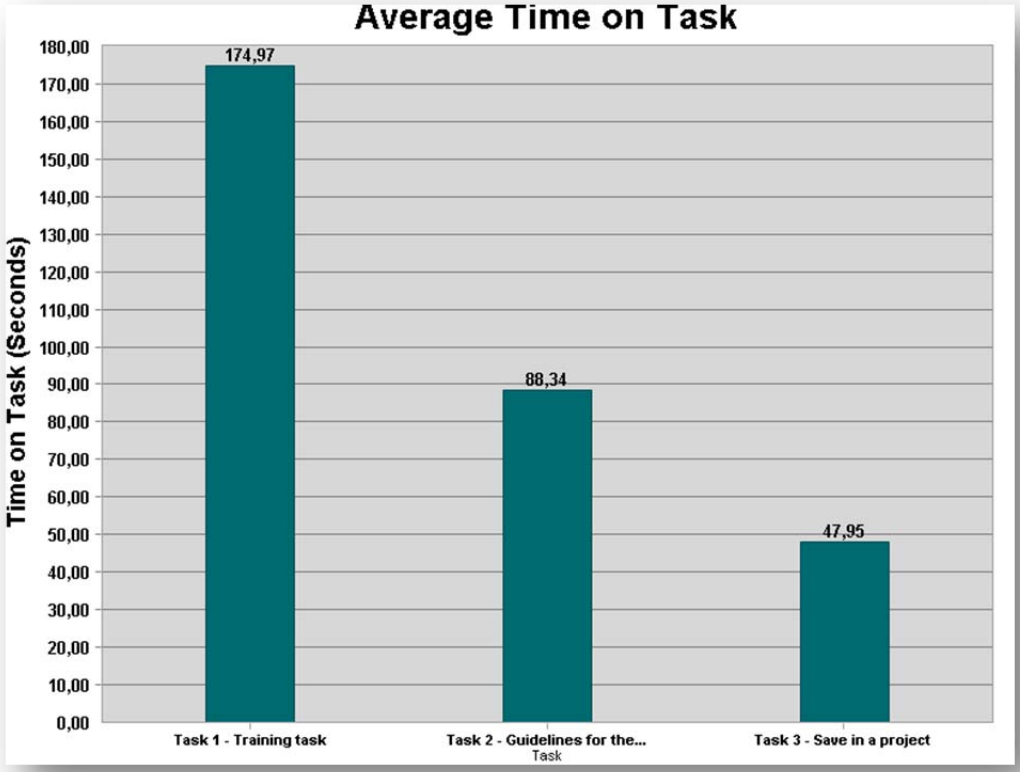


Figure 4.6 Average time on task (to version-1

As Figure 4.6 and Figure 4.8 represent, users decreased the need time to carry out the same first three tasks of both versions. Therefore, although in both cases the time for getting the most suitable heuristics is very low, the needed time is lower when participants are familiar with the interface and the process.

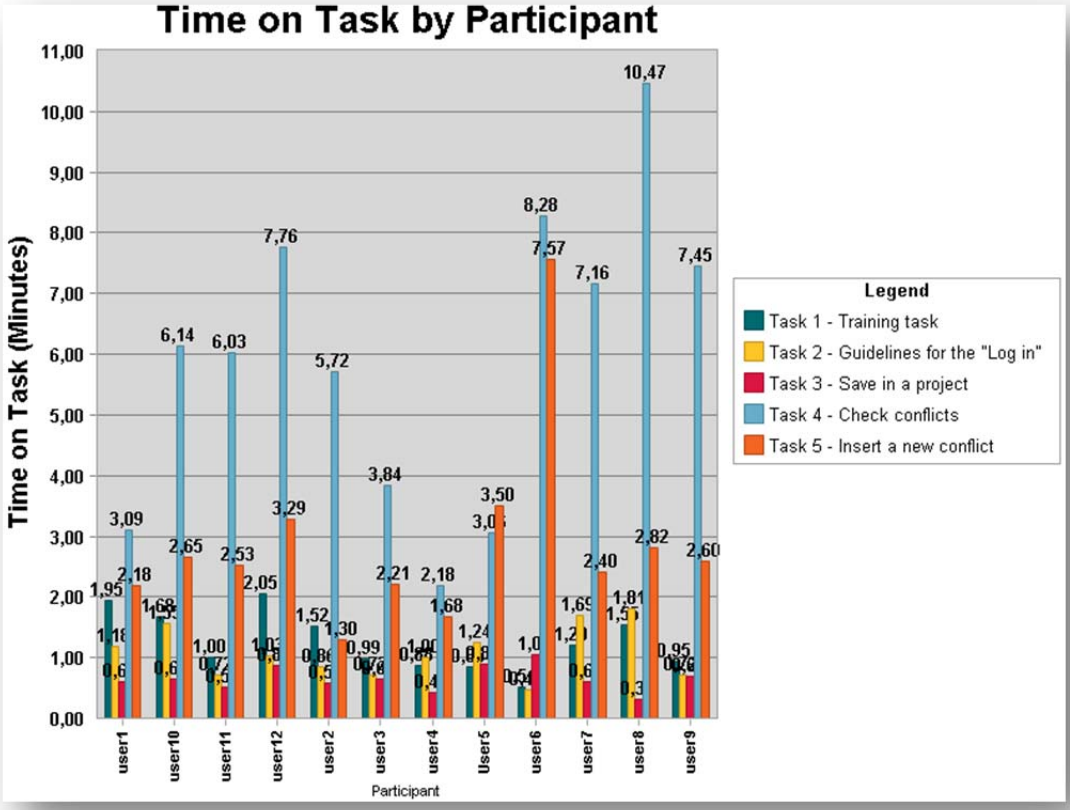


Figure 4.7 Time on task by participant (to version-2)

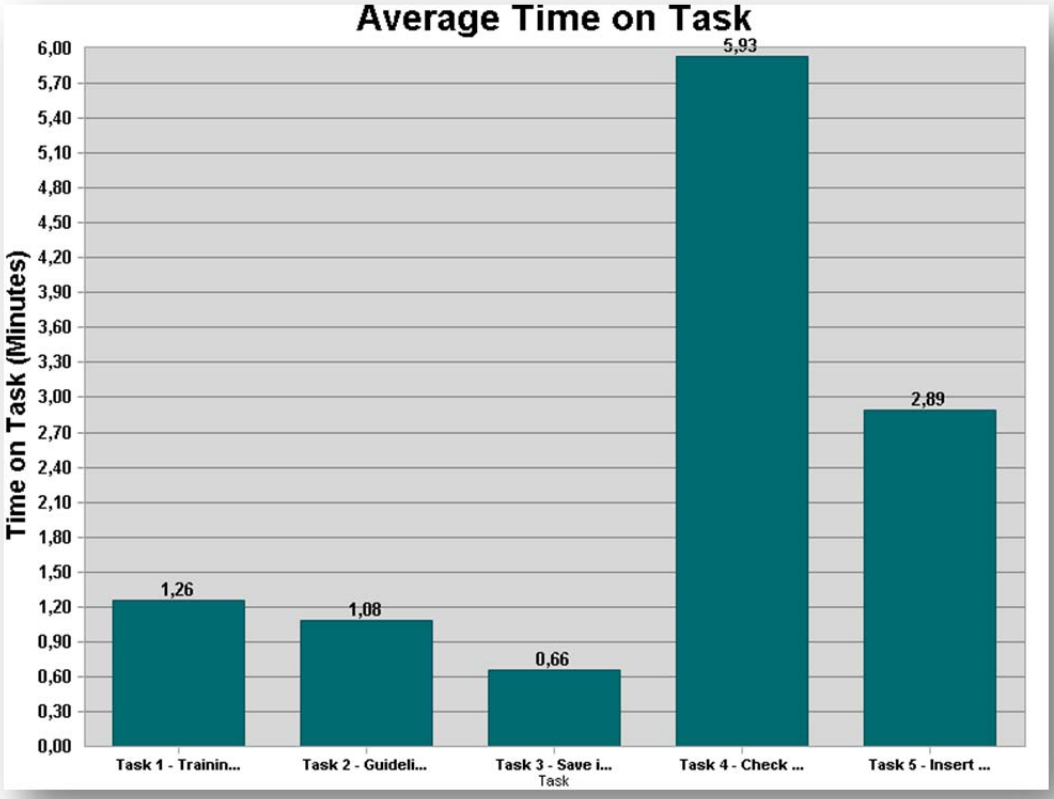


Figure 4.8 Average time on task (to version-2)

4.3.2.1.4 Hypotheses verification

Considering the answers from the pre-test questionnaire about their own experiences using heuristics and the post-task and post-test forms, it is possible to present the hypotheses verification. It is noteworthy that participants do not have knowledge about managing conflicting heuristics and TEAM notation.

- H1. If multiple sources of heuristics are used, it is very likely that users will find conflicts among heuristics over time.
  - H1a. If people use multiple sources of heuristics in their work, they might have found conflicts among them in the past.

This hypothesis is verified through the pre-test questionnaire.

About their experience with heuristics, 5 participants use heuristics in their job very often. 4 participants occasionally use heuristics and 3 of them use them seldom.

Moreover, in 8 cases their organization requires them to use user interface guidelines. In addition, they mainly use heuristics as a source of advice for selecting design options, as a support to inspect user interfaces or as a source of ideas for the design.

When we ask what kind of documents they look at when they need heuristics, participants answered some different options: they mainly look at style guides (11 participants) and standards (8 participants). However, they also look at design rules (5 participants).

Regarding the most frequent sources of heuristics that participants look at, the answers were very different. Table 4.5 shows the results.

	Scientific journal papers	Conference proceedings	Handbooks	Technical reports	PhD thesis	Tutorials	Blogs or websites
Number participants	5	8	6	6	5	8	9

Table 4.5 The most frequent sources of heuristics

However, in any case all of them use more than one source of guidelines. Nevertheless, only one of them used a tool called Jabref (http://jabref.sourceforge.net/) for the management of heuristics.

Finally, when participants answer if they have found some conflicts among heuristics while they are checking different sources, the answers are: 2 participants find conflicts very often. 5 participants find them occasionally, 1 of them find conflicts seldom. Two participants never found conflicts. Finally, two participants do not know if they have found conflicts when they are using different sources of heuristics.

In the cases where the participants have found some conflicts, they solved the conflict through different ways: applying common sense, selecting the heuristics that they most liked, asking customers or using a random method.

- H1b. If people are using tools (such as Open-HEREDEUX that include multiple sources of heuristics), they might be able to find conflicts among heuristics.

This hypothesis is validated checking the successful execution of Task 2 and the answers of its post-task form. The results are divided into both versions.

Version-1- without conflict management:
Everybody finished the task successfully. In addition, 4 participants did not find any kind of conflicts. 7 participants do not know if there were conflicts and 1 participant did not mind.

In this case, 6 participants would like to take note of the conflicts before they use of the set of guidelines because they prefer to know and solve the conflicts before applying the heuristics. 2 participants say that they would not like to take note of the conflicts. However, one of them prefers the conflict management interface before the option of saving the project. In addition, 4 participants do not know when they would like to take note of them. The test recordings show that participants are disoriented given that they felt that they had to find conflicts but they did not find anything.

It is important to highlight that in this version the list of heuristics do not provide any information about previous conflicted heuristics. Therefore, participants cannot detect previous possible conflicting heuristics.

Version-2- with conflict management:
In this interface, everybody also finished the task successfully. Using this interface, 3 participants found some kind of conflicts. 2 participants did not find conflicts. 6 participants did not know and 1 participant did not mind.

In this case, 8 participants would like to take note of conflicts before they use the set of heuristics because they think that the early detection will help evaluators and designers in the improvement of the interactive system and because it is better to solve these conflicts before using the set of heuristics. 1 participant said that he or she do not would like to take note of conflicts. And 3 participants did not know when they would like to take note of them.

- H2. If people are using tools (such as Open-HEREDEUX that includes multiple sources of heuristics), they will be more efficient in recognizing, understanding and reporting conflicts that are clearly exhibited by the tool.
  - H2a. If there are conflicts among multiple heuristic sources, it will be easier for participants to find conflicts if they are documented at Open-HEREDEUX.

In the same way as the verification of the previous hypothesis, Task 2 and its post-task form present the accomplishment of the hypothesis.

Remember the heading of Task 2: Search heuristics for the evaluation of Log In functionality of the website application.

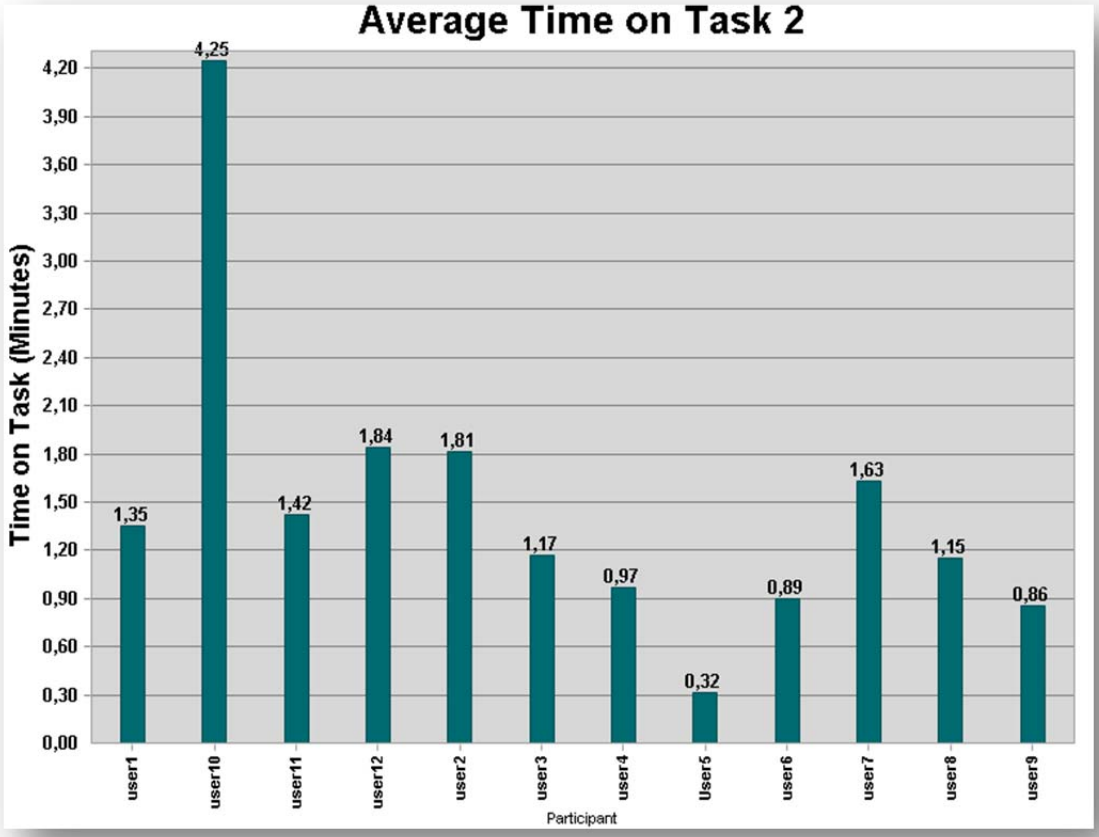


Figure 4.9 Average time on Task2 – version-1- without conflicts

Version-1- without conflict management:

Everybody finished the task successfully. But nobody notes conflicts among heuristics. 4 participants said that they did not note anything. 7 did not know and one participant did not mind.

In Figure 4.9, the average time on task 2 is presented. Although "User 10" spends more than 4 minutes, in all other cases, they do not spend more than 2 minutes. 8 users do not spend more than one minute and a half. Therefore, it is a very efficient task.

Version-2- with conflict management:

In this interface, everybody also finished the task successfully. Using this interface, 3 participants found some kind of conflicts. 2 participants did not find conflicts. 6 participants do not know and 1 participant does not mind.

Those who discover some kind of conflict comment that they see the heuristics marked with an asterisk or that they were reading the heuristics and detecting the conflicts manually. Figure 4.10 shows the average time in task 2 using the interface with the conflict management. In comparison with the interface without the conflict management, only 3 users spend more time in this second interface than in the first one. In any case, the time is less than 2 minutes. Therefore, the interface is very efficient and 9 users get the most suitable set of heuristics in less than one minute and a half.

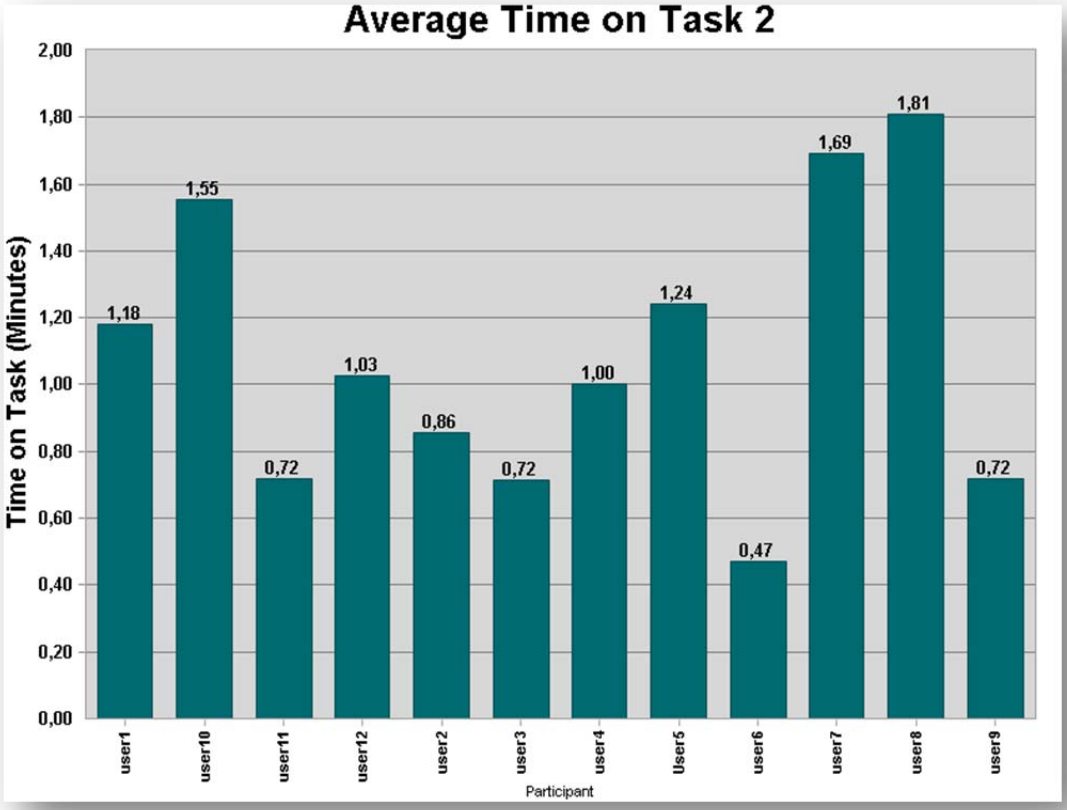


Figure 4.10 Average time on Task2 – version-2- with conflicts

- H2b. If conflicts between guidelines are represented using rationale notation, people will understand the conflicts without additional training on the notation.

This hypothesis is verified with task 4 and its post-task form. This task is only carried out in the interface with conflict management.

At the beginning participants do not understand the TEAM graphic that represents the conflict. Then, although 2 participants decided that they did not understand the graph, the other participants can understand the graph through a short and quick explanation. The main comments of those who understand the graph were:

- "*I can easily understand how to interpret the information with a short explanation from the moderator."*
- *"Different options and the conflict per se are clear but the proposed solution or the solution chosen by other experts is not really clear. This option will need to be highlighted."*
- *"Additional information about the meaning diagram components would be helpful"*
- *"It would be easier if the previous solution is clearly remarked with a stronger line."*
- *"It shows criteria that intervene in each conflict solution."*
- *"The different options I can choose in order to solve the conflict."*
- *"I need some help to understand because I didn't see the solution. Underlining the solution of previous documented conflicts would be interesting."*

Moreover, 6 participants think that the graph perfectly describes the conflicts. 3 decided that it does not represent the conflict because they would like to highlight the chosen solution. Finally, although 9 of them think that this representation is very useful for solving conflicts, 3 other participants do not know because they are not sure about the usefulness of this representation.

Figure 4.11 shows how easy (0) or difficult (5) it is to understand the representation of conflict. Although most the participants think that it is very difficult to understand the graph the first time that they see it, most of them assure that by adding a legend or a brief explanation about how it works, the difficulty decreases enormously. In general participants think that is difficult to understand the representation of conflicts using this rationale notation.

In addition, 6 participants did not change the type of representation but the six others would add a remarked line to clarify the option selected previously for others users.

- H3. If people are using tools (such as Open-HEREDEUX that include multiple sources of heuristics), they might be willing to report their solution or they might understand and trust the solution proposed by others users.
  - H3a. If people found a conflict among heuristics; they are willing to document their solution for conflict resolution.

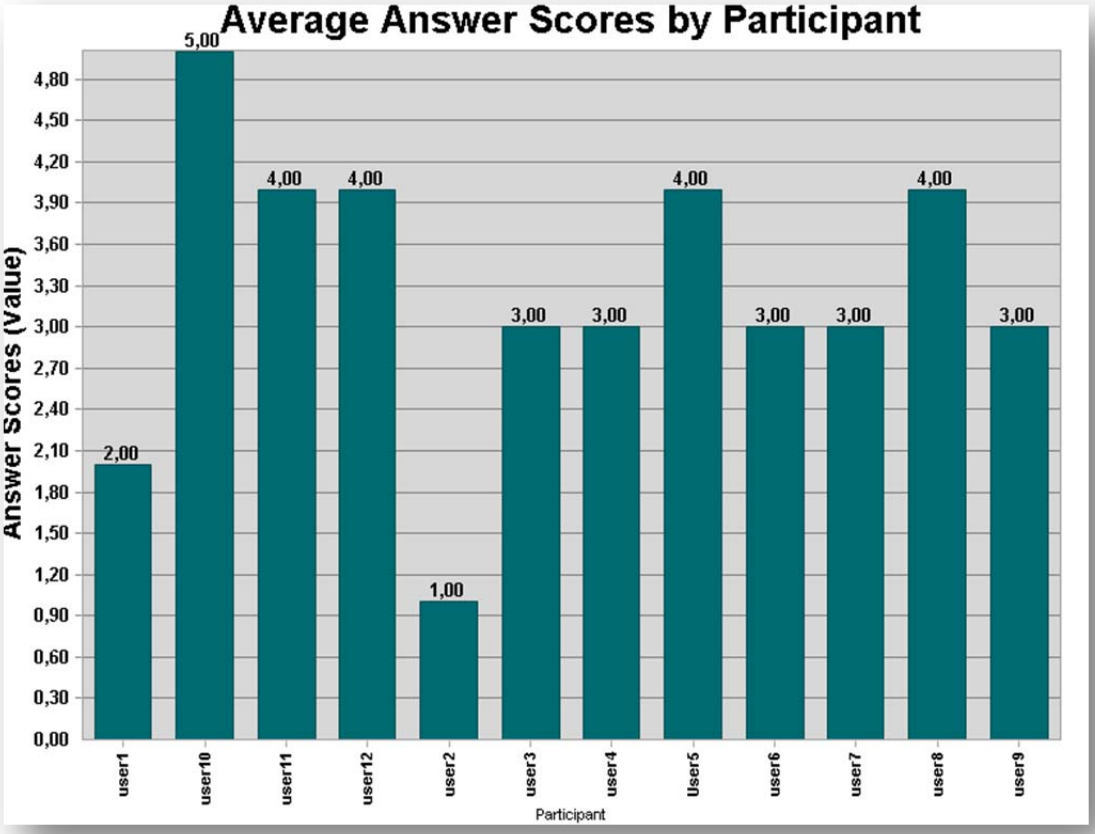


Figure 4.11 Average answer scores by participant about how easy or difficult it is to understand the representation of conflict

Version-1- without conflict management:

According to the post-task 3 form, 6 participants would like to take note of conflicts before using the set of heuristics. 2 participants would not like to take note of them and 4 participants did not know.

Version-2- with conflict management:
According to the post-task 3 form, in this interface, 8 participants would like to take note of conflicts before using the set of heuristics. 1 participant would not like to take note of them and 3 participants did not know.

Somebody who does not prefer the solution using DREAM notation details that they prefer flowchart diagrams or using text and examples.

- H3b. If a tool (such as Open-HEREDEUX) shows solutions for conflicts between heuristics, people would trust those solutions.

Version-2- with conflict management:
11 participants trust in the solution of other people because they suppose that conflicts are documented by experts.

1 participant does not trust it. The participant commented that he or she will consider the resolution but he or she will decide the best one on his or her own.

4.3.2.2 Discussion

This comparative user test validates the usefulness of the version for the conflict management. As is detailed in the hypothesis validation, it is very important to have an interface for documenting the possible conflict that appears when different heuristic sources are used. However, this process should be very clear and understandable for these users.

Although whoever will use Open-HEREDEUX should be experts in UX, a little help is needed for improving the understanding of the conflict management interface (at least the first time that someone interacts with it).

Rationale notation is not usual in the HCI field, so users should have the chance of looking at the help section to be able to use this useful (as they express in the test) option that will facilitate the resolution of some conflicts when they use more than one source of heuristics.

4.4 Scorer of heuristics

Scorer of heuristics (Figure 4.1-3) presents the interface for the realization of the evaluation part of the methodology. It means that it enables carrying out the heuristic evaluation per se. In this component, two types of evaluators are differentiated.

- The first type of **evaluators** is UX experts who will have much knowledge about the UX facets but they might not know the interactive system.
- The second type of evaluators is **end users**, understanding end users as the people who will use the interactive system. Thus, these evaluators can express their emotions better than anybody else despite the fact that they do not have much knowledge about the UX definition and its facets.

For these reasons, I do not say that the resource automates parts of the heuristic evaluation methodology but it automates parts of UX evaluation using a methodology based on heuristics and it helps designers with a list of recommendations that will provide a better UX.

```
In addition, the end user inclusion allows a better UX
evaluation.
```

The task that evaluators and end users can carry out using Scorer of heuristics is the score of the complete set of heuristics provided by Adviser. The scoring process (represented in Figure 4.12) includes the following actions:

- **Fill in two pre-evaluation forms**. As the next component shall detail, results are reported using a template detailed in ISO/IEC 25062:2006. This standard report requires some additional information from the users who carried out the evaluation. This needed information is obtained through two forms. The first one is about personal (age, gender…) and professional information (years of experience in HCI, amount of heuristic evaluations done…). The second form is about tech features such as the browser and operating system, among others.
- **Read the information** about the heuristic: UX degree, UX facet and the description of the heuristic.
- **Choose the specific score** for each severity factor. In any case Scorer uses impact and frequency.
- **Write down some observations** about the heuristics to clarify the score.

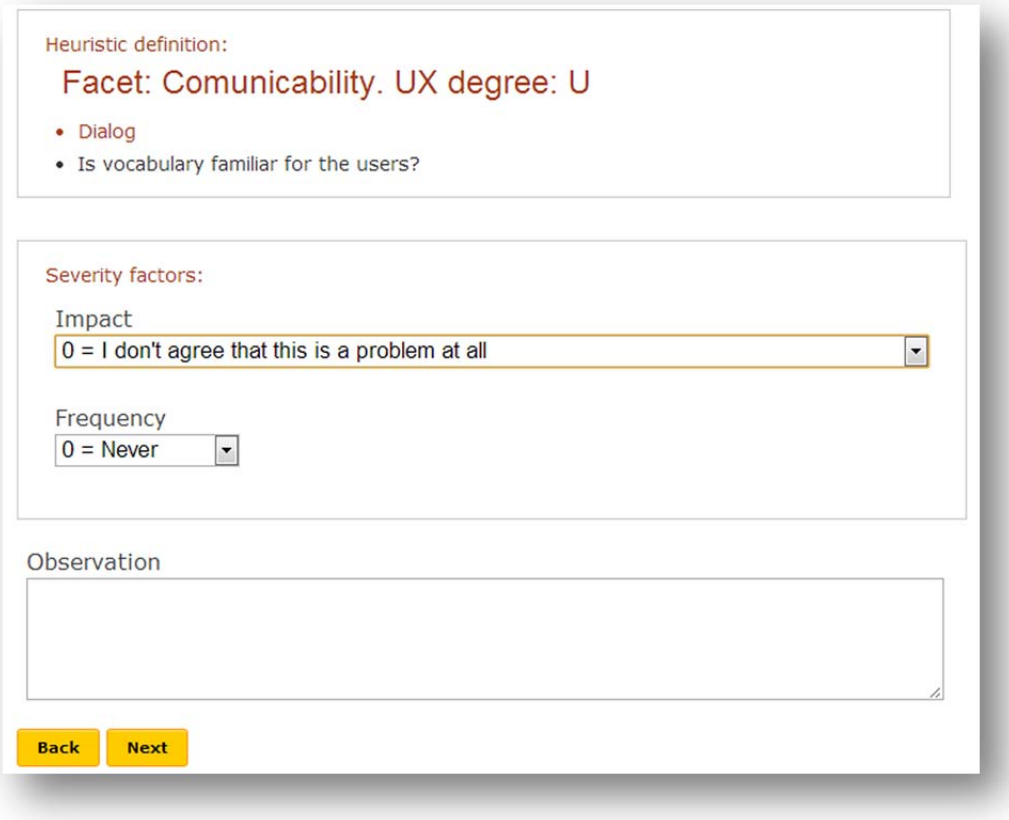


Figure 4.12 Scoring process

The two main advantages when using the Scorer are:

- It **supports the scoring process**. So, the set of advised heuristics can be used directly in Scorer process avoiding manual methods such as paper-based or using any kind of spreadsheet.
- Scorings will determine the **automatic, standard and editable results report** that Result Analyzer will provide.

4.5 Results Analyzer

Results Analyzer (Figure 4.1-4) calculates the results of the evaluation through the set of UX measures presented in the methodology. Therefore, it provides a report with qualitative results and quantitative measures in an editable and standard format file. The most important advantage of this report is the statistical measures that will permit the comparison among different versions of the same interactive system and will enable comparing the results of different systems. This process

could be translated into a certification of the UX evaluation in the near future. Figure 4.13 presents the main page of Result Analyzer.

The main tasks that the project manager can carry out using Results Analyzer are:

- Download the new report.
- Upload the previous downloaded (or not) report for saving it in Open Repository.
- Download the previous uploaded report.

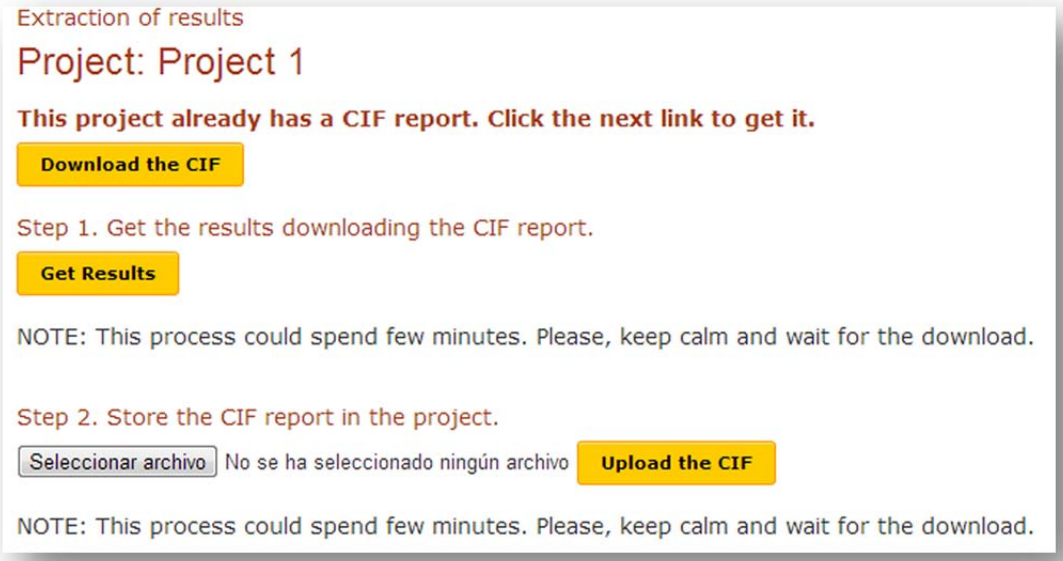


Figure 4.13 Result Analyzer main page

In reference to qualitative results, a list of improvements according to the evaluators' observations and heuristics scored incorrectly will be achieved. They are the ones that designers should apply to the interactive system to improve it.

Apart from qualitative results, quantitative ones should be presented because these will attempt to show the UX degree for every interactive system. If quantitative results are achieved, UX experts will have a standard method or a possible certification to compare evaluations and see which interactive system provides users with the best experience. Quantitative results will get an objective measure that is impossible to obtain due to the subjectivity of qualitative results. This component presents the implementation of the UX measures presented in section

3.5. In the next section a proposal for obtaining UX measures and the editable format to be presented is detailed.

4.5.1 Proposal for acquiring UX measures

Remember that the methodology proposes the next UX measures: CaE (Correlation among Evaluators), automatic classification of problems through clustering methods and PUX (Percentage of UX).

Statistical software will be used to calculate these UX measures. The statistical calculation is carried out through R. R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques (by http://www.r-project.org/). Now, more details about the process for achieving these measures in Open-HEREDEUX using the data from the evaluation and R are described.

> It is very important to highlight that the whole process is absolutely **transparent** for the user. The user only has to click on a button called "Get results" for running the entire process and the CIF report in a .docx format is automatically downloaded.

Results Analyzer runs the next four necessary steps to provide the UX measures (when the project manager clicks on the button "Get Results"):

Step 1. Export data: At this step, the needed data is exported to a .csv file through the Results Analyzer interface. Data is exported following the template from Table 4.6. And in Table 4.7 a real example of the exported data is presented. This process is run by the web server to be completely transparent for the end user.

HeuristicId	SeverityFactor	Mark1	Mark2	…	Markn
H1	Impact	[0-4]	[0-4]	…	[0-4]
	Frequency	[0-4]	[0-4]	…	[0-4]
H2	Impact	[0-4]	[0-4]	…	[0-4]
	Frequency	[0-4]	[0-4]	…	[0-4]
…	…	…	…	…	…
Hn	Impact	[0-4]	[0-4]	…	[0-4]
	Frequency	[0-4]	[0-4]	…	[0-4]

Table 4.6 Template for the results exportation

HeuristicId	SeverityFactor	Mark1	Mark2	Mark3
34	Impact	1	1	2
	Frequency	3	3	1
35	Impact	3	4	3
	Frequency	2	2	2
36	Impact	1	1	1
	Frequency	4	4	4

Table 4.7 Real example for the results exportation

Step 2: Run R code. After creating the file with the needed data, it is time to run the R code. The code is divided into main blocks that use basic statistic functions.

The first block of the R code reads the .csv file to get data and know the amount of evaluators for the evaluation. Bear in mind that the amount of evaluators for the evaluation is not fixed, so it is important to know at the beginning the exact number of evaluators for the specific project. The number of evaluators is the number of columns of the .csv file titled as "Mark#" (where # is a number) (See Table 4.6).

Then the CaE is calculated. Knowing how the Pearson's correlation works, the CaE is calculated doing combinations without repetitions of the number of evaluators. Therefore, in each evaluation there is the number of combination results from the next formula:

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Where "n" is the number of evaluators considered in couples (k=2).

Therefore, the results of this block are the correlation index for each couple of evaluators exported in a text file and a graphic displaying this information. The text file shows the following information considering that the evaluation was done by 3 evaluators:

```
Pearson's correlation coefficient
Correlation between 1 & 2:  0.8279621
Correlation between 1 & 3:  0.8458255
Correlation between 2 & 3:  0.8448457
```

According to the specification of the correlation presented in Chapter 3, in this case, the 3 evaluators present a high correlation. It means that all of them have scored heuristics very similar. And the graph obtained is like Figure 4.14:

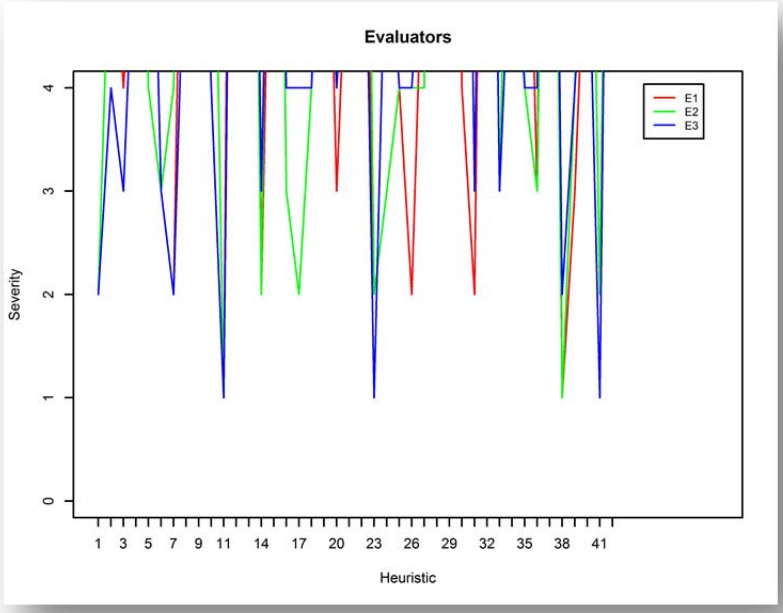


Figure 4.14 CaE Graph

The graph shows details about the specific values of the evaluators' scorings. If the lines are going in the same direction, this means that CaE is high and the evaluators score the same heuristics very similarly. On the contrary, if lines do not appear following the same tendency, this means that CaE is low and evaluators score the heuristic with a very different mark.

The next part of the R code calculates the automatic classification of heuristics using Ward Hierarchical Clustering methodology. Then, the corresponding classification graph is generated and a text file with the heuristics for each gravity level is created. A simple example of this file is presented in the following lines:

```
GROUPS: 1 2 3 2 4
HEURISTICS: 17 32 9 37 19
Average Group 1:  0.3333333
Average Group 2:  1.333333
Average Group 3:  3
Average Group 4:  5.333333
```

Table 4.8 presents an interpretation of these results:

GROUP	1	2	3	2	4
HEURISTIC ID	17	32	9	37	19

Table 4.8 Real example for the automatic classification of heuristics

Then, a lower average of a group means that the gravity level is also lower. Table 4.9 shows the equivalence in the example.

Average Group	0.333	1.333	3	5.333
Gravity level	Minor	Medium	Serious	Catastrophic

Table 4.9 Gravity level equivalence

Figure 4.15 shows this graph. Each (red) square merges a gravity level. Therefore, at this point, the set of heuristics that each gravity level has is exported to a text file. It will be used to show qualitative results about the importance of the problem.

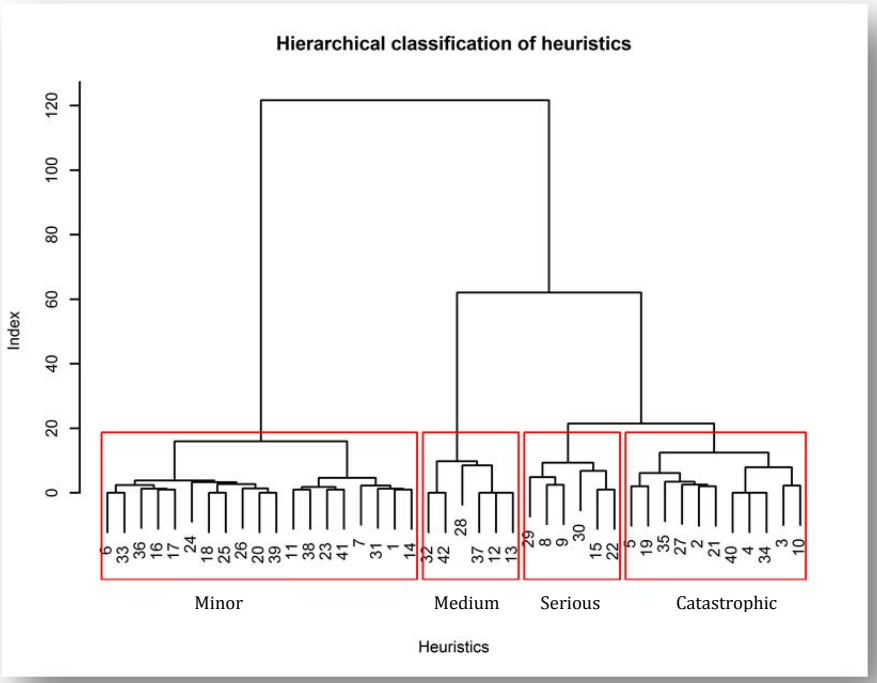


Figure 4.15 Graph for gravity levels

To sum up, after this second step, the results are: the text file with the correlation indexes, the CaE graph, the text file with the heuristics of each gravity level and the graph of the hierarchical classification.

Step 3: Import data. In this step the basic information is saved in the database of Open-HEREDEUX. Therefore, this information is saved automatically to the database to be reached in the next step.

Step 4: Get the results. The last step of the process is the downloading of the final report with the results. Results Analyzer provides users with a file-template based on the ISO/IEC 25062:2006 standard: *Software engineering- Software product quality requirements and evaluation (SQUARE) - Common Industry Format (CIF) for usability reports* [ISO25062]. This template has been adapted for auto-including as much information as possible from Results Analyzer module. With this action, the document provided by the tool follows the CIF standard. Moreover, as the document delivered is editable, it can be freely adapted to the user needs.

4.5.2 CIF report for the evaluation results

The ISO/IEC 25062:2006 [ISO25062] standard presents a specific template to report usability tests based on the types of information that are captured concerning testing with users. The main aim of the CIF is to replace the proprietary formats used by companies. In addition, it presents these main advantages:

- A reduction in training time for UX practitioners.
- An increase of communication between vendors and purchasing organizations sharing a common language and expectations.
- A comparable report with other ones.
- A style guide for non-expert organization that would like to present its reports in a standard way.
- A useful template that can be used in the results of all different and specific cases.

Although the CIF template for usability test reports is the best contribution to standardize these kinds of reports, it presents two challenges due to its applicability in real context:

- The first challenge is the **applicability** of these kinds of reports **in more facets than usability**. It is well-known that usability is the most considered facet in the development process of an interactive system, but up to now other facets have been appearing. Therefore, we need to consider in the CIF the possibility of adapting the results of a test according to other facets such as security, emotional, cross-cultural, among other.
- The second challenge is in reference to the **summative behaviour of the CIF report**. As is explained in [ISO25062] the information that the CIF shows is summative because it tries to present a summary of the results. However, these days, and bearing in mind the competitive sector, much more information should be provided to reach more details about the identification of problems. Therefore, the CIF should include formative information.

Due to both of these needs, the adaptation of the CIF report is proposed. The main differences between the standard template and the proposal are the consideration of the UX as a set of facets in the test and the adaptation of this report in a methodology based on heuristics.

Certainly, there are few publications about how to customize, adapt or modify the CIF for a specific real context. The National Institute of Standards and Technology

[NAT07] changed the CIF for voting systems. Another customization done by the same institute is in the electronic health sector [SCH10].

Furthermore, a research to adapt the CIF to the heuristic evaluation methodology is presented in [VON04]. Using the heuristics evaluation methodology some formative results appear in the CIF report. However, the most informative report is the one presented in [FOR05].

Therefore, according to the last two references, this section presents the customization of the CIF report for the methodology here presented.

Therefore, in Table 4.10 the current sections of the CIF template are presented and checked with those updated in our proposal.

It is noteworthy that some text is presented in all the sections of our template to provide the project manager with the needed information for getting an easier and faster complete report. Therefore, the complete document of our CIF report template is attached in **Appendix C**. In the next subsection, the details about the specific modification of each section are described.

CIF section		**Is it updated?**
Executive summary		No
Introduction	Full product description	No
	Test objectives	No
Method	Participants	Yes
	Context of product use in the test	Yes
	Participant's computing environment	No
	Test administrator tools	No
	Experimental design	Yes
	Usability metrics	Yes
Results	Data analysis	Yes
	Presentation of results	Yes

Table 4.10 The updated sections of the CIF

4.5.2.1 Method section

The information regarding the process of the evaluation is described. To be more precise, the updated sections of the CIF report here detailed are: participants, context of product use in the test, experimental design and usability metrics.

4.5.2.1.1 Participants

In this section of the CIF template, some information about the test users is required. This information is obtained from the two pre-evaluation questionnaires that each evaluator should fill in before starting the evaluation. Therefore, according to the

two types of evaluators considered in the methodology, two different types of participants are considered in this template.

- **End users**: As we commented on before, if the UX would be considered, the end user should be included in the development process of an interactive system. For this reason, the end user who is not an expert in the methodology based on heuristics is considered in the report. Therefore, the novelty is not in the consideration of the end user in the report but the consideration of the end user in the methodology used for achieving formative results.
- **Expert users:** Expert users are those users who know the methodology that is used to evaluate the interactive system. For instance, interface designers, UX researchers or project managers.

The CIF reports this information in two different tables. The table referring to the end users includes information about gender, age, education, computer experience and product or interactive system experience.

The table referring to the expert users presents gender, age, education, computer experience, years in HCI and the amount of evaluations done.

See Appendix C for more details.

4.5.2.1.2 Context of product use in the test

In this section, the Tasks subsection was removed because the methodology based on heuristics does not include any tasks. End users and expert users use the interactive system through a freeway.

4.5.2.1.3 Experimental design

In the same way as the last point and for the same reason, the subsection called Participants task instructions was also removed.

4.5.2.1.4 Usability metrics

First of all, this section has been changed. The proposal calls this section UX metrics. In addition, the metrics for effectiveness, efficiency and satisfaction have been substituted by the definition of CaE, UXD and PUX.

4.5.2.2 Results Section

This section describes how the results of the evaluation are provided.

4.5.2.2.1 Data analysis

Here, the severity factors used to score each heuristics are presented. The methodology based on heuristics uses impact and frequency defined by Nielsen as the severity to score the heuristics. Therefore, the results analysis is based on the scorings of each heuristic per each evaluator using these severity factors.

Then, the CaE measure and the automatic classification of problems are detailed with their respective graphs.

4.5.2.2.2 Presentation of results

Obviously, if the metrics to analyze the UX are changed, the presentation of the results is also modified.

Qualitative results are presented as an improvement list for the interactive system. This improvement list includes the whole set of violated heuristics.

Quantitative results are shown through the PUX of each UX degree and the total PUX. This UX measure compares different versions of the same interactive system and it permits comparison among the same types of interactive systems.

4.6 Benefits of Open-HEREDEUX

Despite some advantages of each component being detailed in previous sections, it is very important to highlight the benefits of the usage of Open-HEREDEUX.

- **A semiautomatic process**: The main advantage of Open-HEREDEUX is the change of a complete manual process, where the best heuristics for a specific interactive system are chosen by the administrator of the evaluation, towards a semiautomatic process, where the administrator only reviews the set of suggested heuristics to approve that the set is sufficient for the specific analysis of the interactive system.

- **A large repository of heuristics**: In addition, taking into account that one of the most difficult tasks is the selection of the best heuristic for the specific system, the first benefit of Open Repository is to provide the scientific and professional community with a large pantry of heuristics (from different reliable sources). The access to a repository where a large set of heuristics is stored (at the moment 505 heuristics) is a very important advantage that facilitates the selection of the best heuristics

by designers and evaluators, instead of searching for different sources and trying to select the best ones manually.

- **Open resource**: Another advantage, which is not less important than the other ones, is the characteristic of open. On the one hand, the open characteristic means that it will be constantly evolving because it is possible to insert/extend/modify the information in an easy and fast way. On the other hand, Open Repository and the whole Open-HEREDEUX is developed using free software and it is also published in a free way because I would like to guarantee that everybody can access the information.

- **For designers and evaluators**: Open-HEREDEUX is used for designers and evaluators of interactive systems that would like to get a system that provides a positive experience. On the one hand, system designers can use Open-HEREDEUX for acquiring design heuristics for specific functionalities, features, components and/or UX facets of an interactive system. These heuristics will be taken as design and useful recommendations for specific interactive systems because they are designed to maximize the user's experience.
  On the other hand, evaluators can use specific heuristics for functionality, feature, component and facet to consider if the evaluated system applies or not to these heuristics. Moreover, Open-HEREDEUX is a good resource to consider in a development course following the basic principles of the user centered design methodology [ABR04] because they can be used in every step of the interactive system development.

- **Automatic and standard result report**: Open-HEREDEUX enables the generation of a standard and automatic result report that provides UX practitioners with a very fast way to get evaluation reports. The most important advantage of this report is the statistical measures presented in previous sections that will permit comparing among different versions of the same interactive system and will enable comparing the results of different systems. This process could be translated into a certification of the UX evaluation in the near future.

- **Individual use**: Finally, although Open-HEREDEUX shares the database with all the resources, Open Repository and Adviser of heuristics work individually.

4.7 Designing Open-HEREDEUX

As experts in the HCI area, Open-HEREDEUX is designed following a user centered design methodology (UCD) [ABR04]. Using this methodology, the end user is included in the development process from the first to the last step of the process. It enables getting a more gratifying product and it causes a more positive experience for end users.

The development process of Open-HEREDEUX includes the next seven steps:

- Analysis of requirements
- Paper prototype
- Focus group
- Software prototype
- First user test
- Redesigning tasks and implementing the entire software prototype
- Feasibility test

4.7.1 Analysis of requirements

The first analysis was the requirement process. The requirement analysis was done with possible end users, people who often carry out usability evaluations using the heuristic evaluation technique. The results of this analysis were the main functionalities of Open Repository and Adviser of heuristics and the main information that the resource should store and manage.

4.7.2 Paper prototype

Following the requirement analysis, a paper prototype was created for showing end users the first idea about how the interface will evolve and how the information will appear in it. The users accepted the prototype with minor changes to be done, so, the database was created to save the information. A MySQL database was defined using an Apache server. According to the ontology presented in section 3.2.7, the needed relations (tables) and attributes (fields) were added to the database to store all the information for Open Repository and Adviser.

4.7.3 Focus group

After including the minor changes proposed in the paper prototype, a focus group was run to detect more improvements and consolidate the database information and relations.

4.7.4 Software prototype

After the database definition, the first software prototype for Open Repository and Adviser of heuristics was implemented.

The implementation was carried out using the PHP language and Cakephp library (http://cakephp.org/). Therefore, Open-HEREDEUX is developed through web technology and, as a consequence, it is available on the Internet for everybody in the world.

It is available at (www.grihotools.udl.cat/openheredeux). Open Repository is open for everybody who would like to look up information. However, the update interface and Adviser of heuristics need a registration to be accessed.

4.7.5 First user test

Once developed the first interactive prototype of Open-HEREDEUX, the first evaluation is carried out. The user test was chosen to evaluate the interaction with Open-HEREDEUX interface.

The main goal of the user test was to evaluate the usability and the main functionalities of the prototype. Bearing in mind that the prototype proposes a new interface for the users, I would like to check if it covers all the required needs to carry out a complete heuristic evaluation.

The test was run in our usability lab (UsabiliLab) and using Morae software to record the needed information.

Due to the nature of the interface, the test should be carried out with users who know the heuristic evaluation methodology because the essential objective of the test was to prepare a heuristic evaluation for a specific interactive system. Now, details about the test are presented.

4.7.5.1 Tasks

The scenario presented to the users was:

"The Alluseful Company is developing a new mobile phone prototype and they would like to check if the new mobile phone provides users with a positive experience. So, as an expert in heuristic evaluation methodology, you assure the Alluseful boss that you will get the list of design and evaluation recommendations to apply in the mobile phone. If these recommendations are applied, the new prototype of the mobile phone will cause a positive experience to the end users who will use this mobile phone.

You are going to use a new resource to get the list of design or evaluation recommendations. This new resource lists the best recommendations or heuristics for a design or evaluation of an interactive system but through a previous configuration of the system. This new resource is called Adviser. "

The test presented seven tasks. The main aim of the test is to check the interaction of the main tasks of the interface. The needed information for each task is given users.

The first task was to get a set of advised heuristics for designing a website (T1). This task enables users to become familiar with the interface.

The following four tasks were the configuration of a new interactive system and its functionalities, components, features and facets. Therefore, the tasks are: to insert a new interactive system (mobile phone) (T2), to create a new functionality (T3), to add heuristics to this new functionality (T4) and to add this functionality and other features, components and UX facets to the new interactive system (T5).

After these five tasks, the next one was to rate each heuristic with its corresponding UX degree (T6). And finally, the user had to get the best heuristics for this new interactive system (T7). The last task was the same than the first one. You will see in the results how the time decreased even though every user uses the interface for the first time.

All users followed the same evaluation process. First of all, each one filled in the pre-test questionnaire to detect his/her user profile. Then, he/she started the tasks. In any case, the needed information to carry out each task was given in the task statement. After each task the post-task questionnaire was answered. Moreover, all users did the tasks in the same order because in most of cases the results of one task

were needed for the next one. Once all users had completed all tasks, they answered the post-test questionnaire or the satisfaction questionnaire.

4.7.5.2 User profiles

Due to the nature of the system, the test should be carried out with users who know the heuristic evaluation methodology because the essential objective of the test for the users was to prepare a heuristic evaluation for a specific interactive system.

Therefore, nine users participated in the test. In Table 4.11, Table 4.12, Table 4.13, Table 4.14 and Table 4.15, different features about the 9 participants are presented (4 men and 5 women). In all cases, they had knowledge about the heuristic evaluation and everybody had done at least 4 heuristic evaluations during his or her entire life.

20-25	25 -35	36 -50
1	5	3

Table 4.11 Ages

Degree	MhD	PhD
2	4	3

Table 4.12 Level of studies

Less than 10	Between 10 and 50	More than 50
4	3	2

Table 4.13 Number of heuristic evaluation done

Nothing	MS Excel
8	1

Table 4.14 Software to prepare the heuristic evaluation

Nothing	MS Excel
6	3

Table 4.15 Software to carry out the heuristic evaluation

4.7.5.3 Results

As a summary, in Figure 4.16, the average time spent in each task is presented. As you can see, the first task was done in 5.12 minutes, but the other ones were only finished in no more than 3.07 minutes. Therefore, I can consider that this time is

very efficient according to the time spent to carry out the same process but manually.

Moreover, according to the task definitions, the first and the last tasks are very similar. The former is to get the best guidelines for a designer and the latter is to get the best set of heuristics for the evaluator. Although the latter has more steps (as you can see in the Adviser process in section 4.3), the user did the task in half the time of the first task. In the first task users spend an average of 5.12 minutes and in the last task (it is noteworthy that this last task is longer (one more step)), only 2.37 minutes were necessary because the user was already familiar with the system. Therefore, after a good configuration of a specific interactive system, the average time for getting the best set of heuristic is 2.37 minutes.

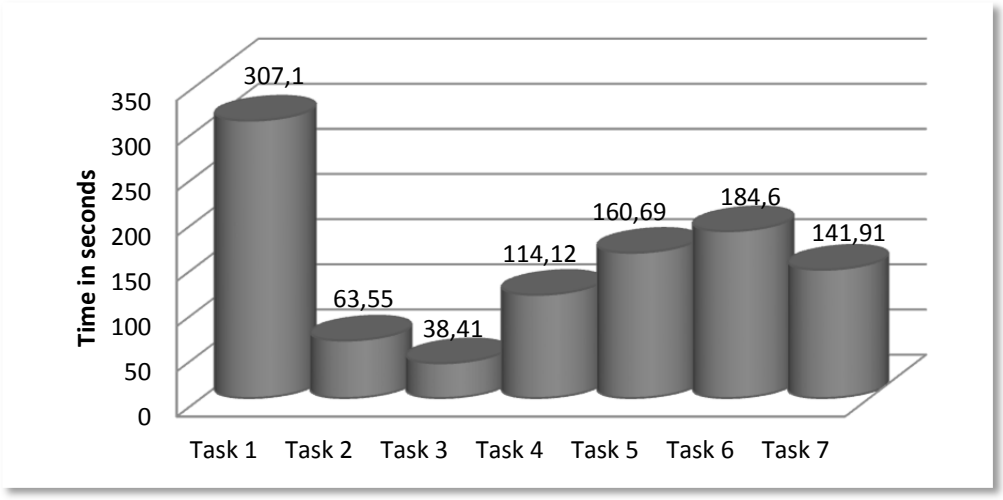


Figure 4.16 Average time in each task

After each task, users answered a post task questionnaire to provide the results with knowledge about the user satisfaction. Users must rate each question with a value between 1 and 4 (both included). Each post task questionnaire has four or five questions. In any case, the first four questions are the same in all tasks. These are:

- Q1: Has the task been easy to do?
- Q2: Has the interface helped you to perform the task?
- Q3: Did you feel comfortable while you were performing the task?
- Q4: Could you repeat the task without any difficulty?

The fifth question, if it exists, is specific to each task. As shown in Figure 4.17 and Figure 4. 18 , the minimum average of scoring of the first question appears in the

first task with a 2.78. But if this task is ruled out because it is the first one and the user is not familiar with the interface, the next minimum scoring is 3.33 of 4.

Except the fifth question from task 2, the other ones exceed the needed scoring (2.5) to consider that the task pass the test according to the specific aspects of the questions.

In the second task, users only create an interactive system according to its definition (the name and the description of the interactive system) and the fifth question asks if users consider that the interactive system definition needs more information. Some users pointed out the possibility of adding more characteristics from the system such as components, among others. They gave these answers because they did not know that they would add these features in the following tasks. In a general view, these answers prove the need of the interactive system parameterization through these different attributes.

Apart from these questions, the post task 4 questionnaire asks, in the fifth question, for the logical presentation of the information according to the structure presented in the interface: First of all users choose the UX facet and after this, the heuristics of the selected facet. Users thought it was the best way to list the heuristics.

Bearing the fifth question for task 2 in mind, the last question for task 5 asks for the other attributes apart from the functionalities, features components and facets to define a specific interactive system. In this question everybody agreed that more aspects are not necessary, the considered information is sufficient for defining a system.

In task 6, the last question asks for the suitability of the UX degree for the heuristics classifications. Every user said it was a good idea to classify the heuristics and to get results in an easier way.

Finally, the fifth question for task 7 concerns the time needed to score two heuristics. According to previous research, two heuristics can be scored in a minute and users have some doubts about it because it is a subjective consideration. But everyone considered that is a good mark of time for starting to consider it in the heuristic evaluations.

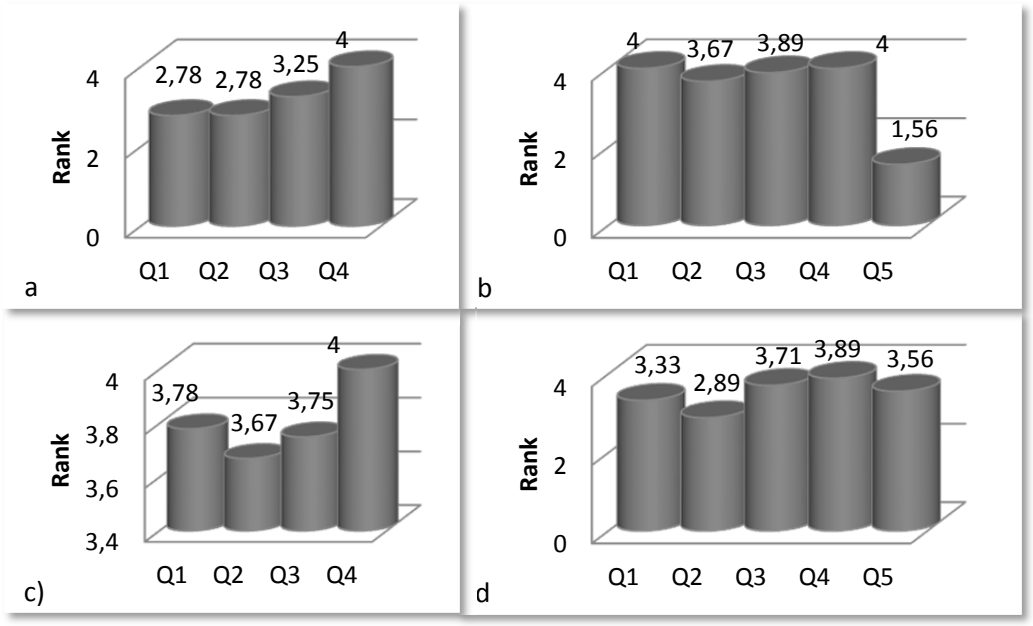


Figure 4.17 Rank of scorings: (a) Task 1; (b) Task 2; (c) Task 3 (d) Task 4

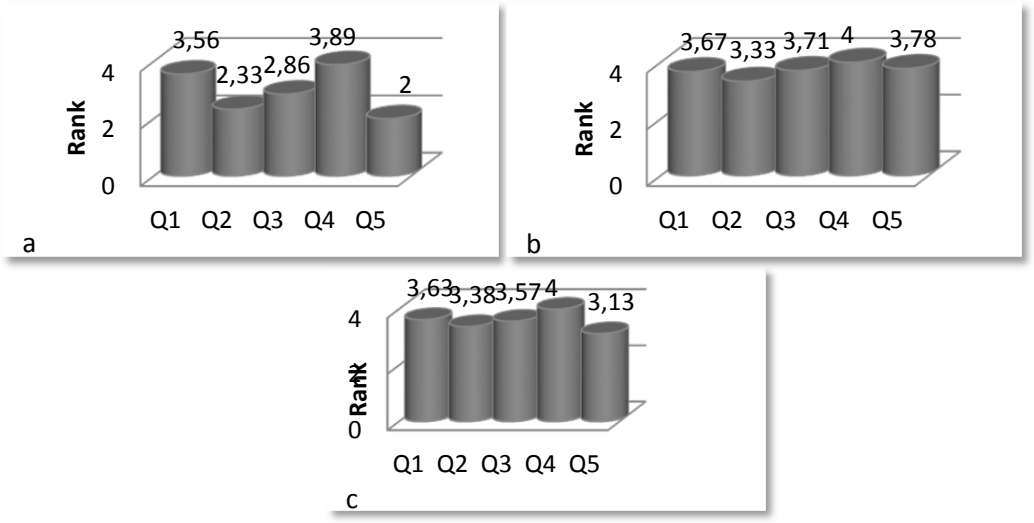


Figure 4. 18 Rank of scorings: (a) Task 5; (b) Task 6; (c) Task 7

Apart from these results, a satisfaction questionnaire was answered by each participant after the test. The answer of the question about if the user thinks the tool is useful, 7 of 9 users said that the tool is useful. And 2 of 7 think that it is a little bit useful because the interface aspect should be improved.

The second question was about the efficiency that the tool has to get the best set of heuristics in relation to the manual process. Everybody says that the tool is better for getting the best set of heuristics.

The next question was if they would recommend the tool to other heuristic evaluation experts. The answer was definitely yes.

The positive aspects mentioned about the tool are:

- To search heuristics in the literature is not needed. This means that we save time because this part is one of the most expensive tasks of the heuristic evaluation.
- It is easy to use.
- This tool is a great help in the evaluations. Moreover, the creation of a wide and complete database improves UX evaluation.
- It can be used in real contexts.
- The flexibility of the repository according to the insertion and modification of the information.
- The choice of different goals to get the best heuristics.
- It is intuitive.

Finally, the last question was about some negative aspects that should be improved. Users answered:

- It is difficult to understand how it is structured the first time.
- The aesthetic design should be improved.
- In addition, 5 of 9 users told us that they will extend our tool including the interface to score the heuristics and to get the results.
- And 4 of 9 participants (maybe because they knew our research) said that they did not include anything else in our resource.

In summary, the users were very satisfied with the system but some aesthetics improvements should be implemented to get a more positive user experience. For us, the most gratifying comments were when users asked for aspects that we have had in mind to include in future versions.

4.7.6 Redesigning tasks and implementing the entire software

At this time, our perspectives about the software prototype were reached. The basic interface for Open Repository and Adviser were well accepted by end users in spite of the proposal of some well justified improvements.

Therefore, these improvements were implemented in the interface. Then, Scorer of heuristics and Results Analyzer were also coded.

After all this time programming the complete interface to get Open-HEREDEUX, a new user test is needed to know the new UX using the complete system.

4.7.7 Feasibility test

The second user test was carried out in a real context of use: an international company was interested in our resource and decided to use Open-HEREDEUX to validate its feasibility in their daily workflow.

Therefore, our methodology implemented in Open-HEREDEUX tool was presented to the international company called GFT Software Factory Iberia S.L.U (www.gft.com). GFT specializes in designing and implementing IT solutions for the financial services industry. The GFT group has 22 offices in seven countries and one of the most challenging departments is the UX department. Its main offices in Spain are in Sant Cugat del Vallès (Barcelona), Lleida and Madrid. So, the UX department was the guest to validate the methodology here presented to obtain UX measures through a methodology based on heuristics.

This real experience in an international company represents a collaboration exercise between a university and a company with beneficial goals for both parts. On the one hand, GFT can appreciate if a tool as Open-HEREDEUX, that implements a specific methodology, improves the design and evaluation of the user interface of their products. On the other hand, I (as a member of the University of Lleida) have real experimental data to check the feasibility of the methodology and the tool here proposed.

Therefore, the research questions to be validated in this user test were:

- Open Repository provides users with all the needed information to reduce the time spent in the selection of the most suitable heuristics for a specific interactive system.

- Adviser solves in an automatic process the main disadvantage of the heuristic evaluation: the manual selection of the most appropriate set of heuristics.
- The editable format of the results report and the usage of the CIF template allow a more dynamic extraction of results in a standard format.
- UX measures are very useful to quantify UX and be able to get, in the future, the standardization of the results.

4.7.7.1 Feasibility test procedure

The test was defined with real users to know the needed improvements to be applied in the methodology and in Open-HEREDEUX resource to get a useful tool for the companies of the same sector.

The procedure for the validation was the free navigation through Open-HEREDEUX to carry out the complete process of the heuristic evaluation methodology. The real users were GFT employers and professors from the University of Lleida.

GFT employers decided to carry out a complete process of the heuristic evaluation methodology, from the planning to the results extraction.

Using the *manager role or project manager role*, GFT started to create a project to evaluate their own corporate website (www.gft.com). They followed the defined process to get it: they added evaluators and severity factors to the project. Then, they used Adviser to get the most suitable set of heuristics for the project.

After the project configuration, 5 employers from GFT and 3 from the University of Lleida (specifically from GRIHO) used Scorer to carry out the evaluation per se.

Finally, the project manager used Results Analyzer to get the report and validate its usefulness and appropriateness in their real business environment.

After ending the process, GFT employers agreed a meeting to discuss the needed improvements in the methodology and in Open-HEREDEUX interface. At the University, professors sent me their contributions.

The test procedure was carried out in an open and general way due to the main goal not being based on knowing specific aspects of the working process or specific features of the resource. The main aim of the test was to get general results to know improvements in our proposed methodology and the usefulness of this tool in a real context of use.

Two different types and related results are described. First, the results for the methodology are shown. Then, the results for Open-HEREDEUX resource are detailed.

4.7.7.2 Results for the methodology

Results of the validations through a set of suggestions are detailed here. Results are divided into positive and negative aspects and improvement for the methodology.

4.7.7.2.1 Positive aspects

The best way to present the positive aspects of the methodology is presenting the comments that GFT employers and professors of the University gave me in literal words. So, some comments were:

- *"It's a very good idea that can really help practitioners of the sector in the evaluation task."(by GFT)*
- *"Getting a final report with statistics of the scorings of all evaluators is infinitely useful." (by GFT)*
- *"Carrying out social heuristic evaluations can revolutionize the UX market. Having a tool that enables the evaluation of my system by many people is truly awesome." (by GFT)*

4.7.7.2.2 Problems

Despite detailing positive aspects, the methodology also presents some problems to be solved.

The first problem appears in the severity factors. As is detailed in other sections, our proposal considers the impact (0, it is not a problem – 4, it is a catastrophe) and the frequency (0, never – 4, always) defined by Nielsen [NIE90] because they are the most used factors to consider in a heuristic evaluation methodology.

The users think that the proposed definitions for the impact and the frequency are not understandable. In addition and according to the received comments, both factors are not related to the writing of heuristics. Sometimes users would prefer bivalent answers as Yes/No and then, if the heuristics present a problem, the chance of choosing the impact and the frequency should appear.

Another aspect about severity factors is the need of rewriting some heuristics to facilitate its scoring. This happens in some heuristics where the writing was changed to negative sentence to not affect the results of the evaluation. For example, the heuristic "Are not there terms from a jargon?" is written to answer "It is not a

problem". However, if the same heuristic is written in an affirmative way as "Are there terms from a jargon?", the answer could be "It is not a problem", meaning that there are terms from a jargon.

Regarding the heuristics presentation, some of them are composed of two simple heuristics using a conjunction such as "and". For example, "Is there in the top and the bottom of the page information about where the users are and the last page visited?" It would be necessary to divide this into two different heuristics.

Finally, the comments about the results of the evaluation are about the automatic inclusion of improvements using the heuristics that present a problem in the interactive system. Another aspect about the result reports is its non-flexibility. Users think that it is essential the chance of updating the template or the option of choosing what information of the evaluation they would like to appear in the report.

4.7.7.2.3 Proposed solutions

According to the problems commented above, the following solutions are proposed:

Regarding heuristics writing, a review is needed to improve their understanding.

In future versions of the methodology, severity factors will be chosen for each project. The project manager will choose if he or she prefers either Yes/No answers or the impact and frequency, among possible new options. Then, the results will be adapted because of the severity factors selected for the project managers.

However, regardless of the severity factor selections, the "non-applicable" option for all heuristics will appear.

4.7.7.3 Results for Open-HEREDEUX

The results of Open-HEREDEUX interaction are described here. Results are divided into positive and negative aspects and improvements for Open-HEREDEUX.

4.7.7.3.1 Positive aspects

In the same way than in the positive aspects of the methodology, the literals expressions of the users were:

- *"It is a versatile tool that allows the evaluation of any type of interactive system." (by GFT)*
- *"It facilitates the heuristics selection for the design and evaluation phases of the development process of an interface." (by GRIHO)*

- *"Open-HEREDEUX, even though it could be improved, provides us with a system to carry out the complete process in a transparent way. But the most important aspect is that it is really fast." (by GRIHO)*

4.7.7.3.2 Problems

One of the problems commented by users is about Scorer. Currently, the evaluation interface runs each heuristic of the project individually. It means that after scoring one heuristic you should click on a button to get the next one. Users told us that this way to present the scoring process is slow and boring.

Another problem in the evaluation process is when the evaluator leaves the evaluation. When he or she connects to the interface again to finish the evaluation, Open-HEREDEUX does not provide him any option or functionality to go to the last scored heuristic. The evaluator should navigate through the list of heuristics to find the last one scored.

Another deficiency of Open-HEREDEUX is the functions related to the user connection. In general, the user management should be improved to provide actions as the "Log out".

Furthermore, the feedback that the interface provides should also be improved. The number of heuristics of the project should be shown in Scorer and the amount of heuristics for finishing the evaluation should also be presented during the scoring process. In general, the feedback messages should be improved towards a more thorough and friendly text.

The project managers missed a very curious functionality: the possibility of keeping a watch on the evaluator's tasks. They would like to monitor the current state of the evaluator's work.

Finally, the aesthetic design of the whole resource should be improved.

4.7.7.3.3 Proposed solutions

According to the problems described in the previous sections, the following improvements are proposed:

Regarding Scorer of heuristics, the heuristics will be presented per UX facets, all of them in the same screen. Therefore, the evaluator will only change the screen after the scoring of each UX facet.

In addition, if the evaluator leaves the evaluation, a warning message will appear when the evaluator comes back and the chance to go directly to the UX facet and the

set of heuristics will be presented.

An important effort will be done in the user management. Currently, the user management includes a basic management such as the creation of different user roles and each role has some permissions. In future versions of Open-HEREDEUX the creation of a list of users will be managed by project managers. This new functionality will allow project managers the management of their own users for each project.

The feedback will be enhanced definitely, adding status information in Scorer and improving the general text presented in messages.

In general, the style sheet will be updated to get a more beautiful interface.

Finally, two new modules will be presented. One module will enable project managers to review the evaluator's scorings. Another module will allow more flexibility for obtaining results. Project managers will see the real state of the evaluations (those who just finished the evaluation and those who have not) and he or she will be able to decide what evaluation will take part of the results, enabling the chance of omitting some of them or some information of the evaluation.

4.7.7.4 UX measures opinions

The most important goal in the definition of this set of UX measures is the acceptation of these measures in a real context of use, in other words, in a business environment. This means that an international company would consider that CaE, PUX, UXD and the automatic classification are useful for its daily workflow.

3 employees from GFT Company answered the questionnaire. The results are detailed following each question.

- Do you think that the format of the document (.docx) is correct due to the needed use of the information?

Two of three answered that the format is fine because they can edit the document. The third thinks that .doc is better because it is still more common.

- What do you think about the use of the ISO/IEC 25062 Common Industry Format for usability test report to present the results of the evaluation?

GFT employees think that it is the most appropriate format. The format used is OK and the structure of contents as well. Even so, it looks very boring and they wish there was more highlighted text in bold for making reading easier. They would prefer the same content/info in a different doc template (a more creative template using colored fonts in titles for instance).

- Do you think that the correlation among evaluators is useful and substantial for the results of the evaluation?

Everyone thinks it is useful and substantial and they added some comments: its explanation must be widely extended in the document and graphical information is difficult to interpret. In addition, the information in the chart needs to be explained for easy interpretation.

They also added that it is very important because this correlation between evaluators can ease the clustering of different intuitive judgments, ergo different mental models.

- Do you think that the automatic classification of the heuristics is useful and substantial for the results of the evaluation?

Everyone said that it is useful because it facilitates the evaluation work grouping errors under the same heuristic typology for making the problem-solving process easier. The structure of categories is always useful as it helps to group related concepts.

- Do you think that the percentage of UX is useful and substantial for the results of the evaluation?

Everyone selected: Yes.

- Who can be the receiver of the CIF report?

GFT employers decided that this report is useful for project managers, designers, developers and client.

4.8 Conclusions

One of my personal goals when I started my thesis was not to finish only with theory, but to develop a tool for supporting the methodology. Open-HEREDEUX is the resource that implements the methodology based on heuristics (detailed in chapter 3). Apart from the advantages detailed in section 4.6, Open-HEREDEUX provides companies with an added value to facilitate the consideration of the UX in the development process of an interactive system.

The validation of the resource in a real context of use thanks to the GFT Company gives me the necessary validation to prove that, even some interaction improvements needed, Open-HEREDEUX is a useful proposal for companies to test UX projects.

Chapter 5

"The more you have, the more you want"
"Com més tens, més vols"

Conclusions and future research lines

5.1 Conclusions

The main goal of this PhD is the definition of a quick and cheap methodology that permits the consideration of UX in different steps of the development process of an interactive system.

To reach this goal an exhaustive study of the UX concept was needed to exactly know what it means and includes. In addition, evaluation methodology with a contrasted efficiency together with quality standards in this area were the basis of this research.

Therefore, a semi-automatic methodology based on heuristics for considering the UX in the design and evaluation steps is proposed. In addition, this methodology is materialized in a technological resource that provides HCI community with a real tool to be used.

L'objectiu general d'aquesta tesi és la definició d'una metodologia ràpida i econòmica que permeti considerar l'experiència d'usuari (UX) en diferents fases del procés de desenvolupament d'un sistema interactiu.

Per assolir-ho ha estat necessari estudiar a fons el propi concepte de la UX per tal de conèixer exactament a què es fa referència quan s'utilitza aquest terme. Tanmateix, ha calgut partir d'una metodologia d'avaluació d'efectivitat contrastada així com tenir presents els referents de qualitat establerts en l'àmbit d'estudi.

Així doncs, s'ha proposat una metodologia semi-automàtica basada en heurístiques que considera la UX en la fase de disseny i avaluació d'un sistema interactiu i que es materialitza amb un recurs tecnològic que fa possible l'ús de la metodologia a tercers.

This new methodology is based on the very well-known methodology called heuristic evaluation. However, this new methodology improves the following steps: the selection of the most suitable heuristics for each specific case, the scoring of heuristics and the extraction of results.

Aquesta nova metodologia es basa en la ja molt coneguda i utilitzada metodologia de l'avaluació heurística però millorant les següents fases: l'elecció de les heurístiques més adequades per cada cas concret, la puntuació d'heurístiques i l'extracció de resultats.

First of all, it improves the evaluation planning. Specifically, the selection of the most suitable heuristics for a specific system is semi-automated. This part of the methodology is one of the most undefined because information is dispersed around the different sources in the literature. These suitable heuristics are reached through the creation of a repository of information (modeled through an ontology) where information is stored and connected to carry out a complete heuristic evaluation. This information is: heuristics, facets that define the UX, attributes of the quality standard ISO/IEC 25010:2011, components, functionalities and features that determine interactive systems (mainly websites, virtual assistants, public kiosk and, recently, interactive digital television applications). Obviously, the repository also stores the needed semantic relations that can be used during the semi-automatic process where the most adequate set of heuristics for each specific situation will be provided. This process is called the suggestion of heuristics.

En primer lloc, es millora la fase de planificació semi-automatitzant un dels aspectes que, per la seva indefinició o dispersió, pren més temps com és l'obtenció de les heurístiques més adequades per cada cas concret. Aquestes heurístiques s'obtenen gràcies a la creació prèvia d'un repositori d'informació (modelat mitjançant una ontologia) que emmagatzema i relaciona totes les dades necessàries per realitzar l'avaluació heurística. Aquestes dades són: els criteris heurístics, les facetes que defineixen l'experiència d'usuari, els atributs de l'estàndard de qualitat del software ISO/IEC 25010:2011, i els components, característiques i funcionalitats que determinen com són els sistemes interactius a dissenyar i/o avaluar (principalment webs, assistents virtuals, quioscos públics i, recentment també, televisió digital interactiva). I, evidentment, el repositori també guarda les relacions semàntiques necessàries per ser utilitzades durant el procés semi-automàtic, al que anomenem procés de recomanació, que proporcionarà quin és el conjunt de criteris heurístics més adequat per a cada situació.

Furthermore, UX degree (U, UU, UUU) and the needed time to consider a heuristic in an evaluation are presented as financial constraints with the main goal of reducing the final number of heuristics according to the defined budget in this step of the development process. Both constraints can be applied before the scoring process, thus ensuring the realization of the evaluation according to the available budget.

Another important aspect of the suggestion process is the opportunity to document conflicting heuristics.

Using a specific rationale notation it is possible to represent possible contradictions among heuristics obtained from different sources and for very different application domains.

The detection and documentation of these conflicts is an additional part of the methodology that will improve the use of heuristics in future applications. In addition, it solves possible conflicts among heuristics in a quick and satisfied way.

El grau d'experiència d'usuari de les heurístiques (U, UU, UUU) i el temps que es necessita per avaluar cada heurística tenen com objectiu principal acotar la quantitat d'heurístiques d'acord al pressupost destinat a la fase de disseny i d'avaluació del projecte. Les dues restriccions financeres que es presenten poden ser aplicades al llistat d'heurístiques abans d'incorporar-les al "Scorer" de manera que ja es pugui procedir a realitzar l'avaluació ajustada al pressupost disponible.

Un altre aspecte molt important del procés de recomanació és la possibilitat de documentar conflictes que poden aparèixer entre heurístiques.

Utilitzant una determinada notació formal s'ajuda a representar possibles contradiccions que poden aparèixer quan s'utilitzen heurístiques recollides de diferents fonts d'informació i amb objectius d'aplicació molt diferents.

La detecció i possibilitat de documentació d'aquests conflictes és un complement a la metodologia que s'ha hagut d'afegir per millorar usos futurs de les heurístiques per resoldre les situacions de conflicte ràpidament i de forma satisfactòria.

Secondly, the scoring of heuristics is still presenting in a manual way because the methodology works through the scorings of expert users (HCI experts as well as interactive system experts) and it is impossible to replace these users for machines. However, the needed support to carry out this manual step in an easy, guided and comfortable way is provided. In addition, in this step the methodology stores all needed data to be able to automate the last step of the heuristic evaluation: the extraction of results.

Finally, this last phase, in the same way as the first one, is another important contribution of the PhD. It entails an appreciable improvement of the whole process, mainly in the phase of extraction of results. Up to now, this phase was only subjected to qualitative considerations from experts. However, the proposal contributes new statistical measures for the substantial improvement of the result report interpretation. These measures are:

- The correlation among evaluators to value the reliability of evaluators' scorings.
- The automatic classification of problems enables the removing of post-evaluation meeting where all evaluators should discuss the different scores in the same specific heuristic.

En segon lloc, la fase de puntuació de les heurístiques es continua plantejant segons la seva realització de forma manual, donat que és una metodologia basada en valoracions realitzades per experts (tant si ho són en HCI com si ho són sobre el sistema interactiu en particular) que no poden pas substituir-se per màquines. Es dóna, no obstant, el suport necessari per realitzar-ho de forma fàcil, guiada i còmoda. A més, està previst que es guardin de forma automàtica totes *les dades necessàries de l'avaluació per poder automatitzar la última fase de l'avaluació heurística: l'extracció de resultats.*

Finalment, aquesta última fase, igual que la primera, és un dels punts més rellevants del treball aportat. Comporta una millora considerable del procés sencer, sobretot pel que fa a la seva fase final d'obtenció de resultats. Fins ara aquesta fase estava tant sols subjecta a les consideracions qualitatives dels experts però ara la metodologia aporta noves mesures estadístiques que milloren substancialment la interpretació dels resultats finals. Aquests mesures son:

- *La correlació entre avaluadors per poder valorar la fiabilitat de les puntuacions dels avaluadors.*
- *La classificació automàtica de problemes de manera que s'elimina la típica i costosa reunió post-avaluació, on tots els avaluadors han de discutir aquelles heurístiques que no han puntuat amb els mateixos valors.*

- The UX percentage allows the comparison of results among different versions of the same system or different interactive systems.

The implementation of this methodology is materialized in Open-HEREDEUX (Open HEuristic REsource for the Design and Evaluation of User eXperience). A resource (developed during the PhD period) that enables the realization of UX evaluation based on heuristics. In addition, Open-HEREDEUX provides designers with the needed recommendations for considering UX in the design step of the development process.

Furthermore, Open-HEREDEUX results are presented automatically and using the format presented in the ISO/IEC 25062:2006 standard (*Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Common Industry Format (CIF) for usability test reports*).

As a personal conclusion and regarding Open-HEREDEUX, when I started my PhD, I wanted to make the methodology reality. This means that I do not only want to define a new theoretical method. I wanted the theoretical part (it is essential for getting the next part) but I also wanted to apply this theoretical concepts in a real tool.

- *El percentatge d'experiència d'usuari que permet comparar els resultats entre diferents versions del mateix sistema interactiu i entre diferents sistemes interactius de propòsit similar.*

La implementació d'aquesta metodologia es materialitza amb Open-HEREDEUX (Open HEuristic REsource for the Design and Evaluation of User eXperience), una eina (desenvolupada durant el període de realització de la tesi) que permet realitzar avaluacions de UX basades amb heurístiques i dóna suport al dissenyador llistant les recomanacions de disseny més adequades a cada cas.

També es proposa, mitjançant Open-HEREDEUX, una presentació automàtica dels resultats seguint l'estàndard ISO/IEC 25062:2006 (Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Common Industry Format (CIF) for usability test reports).

Ressaltar que quan vaig començar la tesi, volia fer de la metodologia una realitat. Això significa que no només pretenia definir una nova metodologia teòrica (essencial per aconseguir la següent part) sinó que pretenia aplicar aquests conceptes teòrics a una eina real. Així doncs, el meu objectiu principal i prioritari era aconseguir Open-HEREDEUX com a eina que implementés la metodologia.

So, my personal and priority goal was to make Open-HEREDEUX a tool that implements the methodology. Therefore, Open-HEREDEUX permits the execution of this methodology in real projects. So, considering the definition of the methodology as the main part of the research, the implementation of it becoming a real resource to be used in real projects provides the second most important contribution of this project and for the HCI community.

Currently, Open-HEREDEUX is a software prototype that carries out UX evaluations. Despite being detailed in the results of user tests, some improvements are needed. Furthermore, the feasibility analysis of Open-HEREDEUX in a real company gives added value to the methodology and tool that will make possible future relations.

Apart from this kind of conclusions, it is important to point out a very important aspect of the development process of an interactive system. The main character of this project is the methodology based on heuristics, but this does not mean that the application of this methodology is enough to verify the UX of a system.

Per tant, Open-HEREDEUX permet l'execució de la metodologia en projectes reals. Així doncs, si la part més important de la tesi és la definició de la metodologia, la segona contribució més rellevant és la implementació d'aquesta metodologia a un recurs real que pot ser utilitzat en el context de projectes empresarials existents.

Actualment, Open-HEREDEUX és un prototip software que permet realitzar avaluacions de la UX, tot i que tal i com el test d'usuaris detalla, podria ser millorat. Un altre aspecte important a destacar és que la part experimental de l'eina s'ha realitzat en l'àmbit d'una empresa internacional de producció de software que disposa d'un departament de UX emergent. Per tant, tot i els aspectes millorables, l'anàlisi de viabilitat de Open-HEREDEUX en una empresa multinacional dóna, tant al sistema com a la metodologia, un valor afegit que permetrà futures relacions.

A part d'aquest tipus de conclusions, cal dir que el personatge principal d'aquest projecte es la metodologia basada en heurístiques però això no significa que l'aplicació d'aquesta metodologia sigui suficient per verificar la UX d'un sistema interactiu.

This discount methodology can be applied in different steps of the development process but other methods (for instance user tests) should be considered to find more problems or validate the problems found avoiding some deficiencies of the methodology (such as false positives).

In my opinion, more than one methodology is needed (regardless of the type of methodology) to get an interactive system that provides end users with a positive experience.

L'aplicació d'aquesta metodologia té un preu molt reduït i pot ser aplicada en diferents etapes del procés de desenvolupament d'un sistema interactiu però s'han de considerar altres mètodes (com per exemple un test d'usuaris) per trobar més problemes o validar els detectats i així contrarestar les deficiències de la metodologia basada en heurístiques (com podrien ser els falsos positius).

Com a opinió personal, dir que és necessari aplicar més d'una metodologia (independentment de quines siguin) per aconseguir que un sistema interactiu provoqui una experiència positiva als usuaris que l'estan utilitzant.

5.2 Future research lines

Although we believe that our methodology presents a good basis to start in the consideration of UX in the development process of an interactive system. There are few points that should be worked on in the near future:

- **Upgradeability of the information to enable the design or evaluation of any type of interactive system**. It is obvious that currently our database called Open Repository only stores data for four different case studies: website applications, virtual assistants, iTV and public kiosks. The upgradeability of this information is needed to cover as many different interactive systems as possible to get a valuable tool where the most suitable heuristics for any interactive system can be found.

- **Selection of the most suitable heuristics for each stage of the development process**. In this project, heuristics are proposed to be used in the design and evaluation stage of the development process. However, the consideration of more heuristics in other stages such as the prototyping stage could also be an interesting research line to be considered. Some research questions in this area could be: Which stages of the development cycle are appropriate to present heuristics? Can we use the same heuristics as in the design or evaluation stage? Where do these heuristics appear in the methodology?

- **Automatic process for inserting information, mostly heuristics.** There will be a very useful task. Currently the insertion of new information is a totally manual process that the administrator has to carry out regarding his/her own experience and knowledge about the resource. The automation of this process will allow the insertion of other user roles that perhaps do not know perfectly the internal system of Open-HEREDEUX.

- **Automatic detection of conflicting heuristics**. This improvement will help in two ways. Up to now, the selection of conflicting guidelines is a manual process that permits their documentation to get a set of heuristics that is as accurate as possible. The automatic detection of conflicting heuristics will help project managers in the decision in which heuristics are in conflict through an automatic way and which is the best option to select for a specific context of use. Second, the automatic selection of conflicting heuristics will help the integrity and consistency of the database. If a good implementation of the automatic detection of heuristic is reached, the conflicting heuristics or the duplicated heuristics will be detected when someone inserts more new heuristics in Open Repository.

- **Collaborative methodology based on heuristics**. Although our proposal removes the meeting that proves the collaborative feature in the evaluation, the consideration of this feature not only for the results of the evaluation but for a more general discussion about the interactive system could be researched. A possibility could be the creation of a debate or forum in the specific project to give any kind of interesting information about the UX evaluation of each specific case.

- **More flexible result report**. When we started working on the result report, we believed that the use of an editable format to be able to modify the document was a very good, cheap and fast option. However now, the improvement is on the road to the implementation of the report updates in a web-based resource. In Result Analyzer the option to configure the report and insert all the needed information should appear in the interface to get the PDF report directly in a very good format. It will provide the methodology and Open-HEREDEUX with an added-value to get the report according to specific details, formats and specific presentations of each HCI practitioner who decides to use Open-HEREDEUX.

- **Analysis of evaluators' statistics to choose the best ones in each case**. One of the more difficult and interesting challenges of methodologies that experts need to carry them out is the choice of the best experts in each specific case. According to our result report, it presents some statistical measures that could enable knowing the experience that an evaluator has. Specifically, the correlation among evaluators can help to decide if one evaluator is more experienced in the system and/or the heuristics and/or the tool and it will also help in the choice of one or another evaluator in each specific situation.

- **Setting the most subjective heuristics for the emotional facet.** As this document presents, end users of the applications are included in the evaluation stage. The proposal suggests the most subjective heuristics for this type of evaluators. Although the result report already considers this type of evaluators, these subjective heuristics are not detected. Therefore, an interesting future research line will be the detection of the most subjective heuristics to be able to include end users as evaluators of the interactive system.

- **Different weights for facets and heuristics.** Currently, heuristics are cut in accordance with the order in the database when the time restriction is applied in the set of heuristics. Different weights should be provided for each facet and heuristic to get more precise results in the heuristic suggestion process. These weights can be proposed for a project or if there is a consensus, they could be proposed for a type of interactive system.

- **Cost-benefits analysis**. During the research the topic "discount methodology" appeared. Although I consider that whoever is expert in heuristic evaluation methodology will note that our proposal is faster and cheaper than the traditional and manual one, I should carry out an exhaustive and statistical research for this validation. Our approach should be compared to other inspections methodologies to justify the economic benefits here presented. Furthermore, this analysis should also consider different types of interactive system and number of heuristics to be able to prove the usefulness of the methodology in many different real situations.

5.3 Results and publications

Regardless the repository of information and, in consequence, Open Repository the following communications are published:

Masip, L., Oliva, M., Granollers, T. 2010. **Hacia la semiautomatización de la evaluación heurística: Primer paso, categorización de heurísticas**, *Interacción 2010*, Grupo Editorial Garceta, Valencia, 7 to 10 of September, pp. 359-368.

Oliva, M., Masip, L., Granollers, T. 2010. **Evaluación de usabilidad y accesibilidad de un conjunto de dispositivos interactivos denominados Puntos de Información Ciudadana**. *Scire: representación y organización del conocimiento*, Vol. 16, pp. 35-46.

Masip, L., Granollers, T.,Oliva, M. 2011. **A Heuristic Evaluation Experiment To Validate The New Set Of Usability Heuristics.** *Proceedings of 8th International Conference on Information Technology: New Generations*. Las Vegas. 978-0-7695-4367-3/11 © 2011 IEEE.

Masip, L., Oliva, M., Granollers, T. 2011**. Classification of Interactive System Components Enables Planning Heuristic Evaluation Easier.** *Design, User Experience, and Usability. Theory, Methods, Tools and Practice. Lecture Notes in Computer Science.* Marcus, A. (Ed.). Springer Berlin / Heidelberg. Vol. 6770, pp 478-486. ISBN: 978-3-642-21707-4.

Masip, L., Oliva, M., Granollers, T. 2011. **OPEN-HEREDEUX: open heuristic resource for designing and evaluating user experience**. *In Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction* (INTERACT'11), Pedro Campos, Nuno Nunes, Nicholas Graham, Joaquim Jorge, and Philippe Palanque (Eds.), Part IV. Springer-Verlag, Berlin, Heidelberg, pp. 418-421.

Masip, L., Oliva, M., García, R., Granollers, T. 2012. **Towards Usability Improvement of Semantic Web Applications.** *WEBIST 2012*.Oporto, pp 361-366.

Masip, L., Oliva, M., Granollers, T. 2012. **The open repository of heuristics**. *In Proceedings of the 13th International Conference on Interacción Persona-Ordenador* (INTERACCION '12). ACM, New York, NY, USA, Article 4 , 2 pages.

Solano, A.**,** Masip, L., Granollers, T., Collazos, C.A., Rusu, C. **Setting usability iTV heuristics in Open-HEREDEUX**. Accepted paper in 6th Latin American Conference on Human Computer Interaction (CLIHC 2013).

For the Suggestion of heuristics part of the methodology and the implementation in the Adviser:

Masip, L., Oliva, M., Granollers, T. 2011. **User Experience Specification through Quality Attributes.** *In Proceedings of the 13th IFIP TC 13 international* conference *on Human-computer interaction*. INTERACT'11. Pedro Campos, Nuno Nunes, Nicholas Graham, Joaquim Jorge, Philippe Palanque and Marco Winckler (Eds.), Part IV. Springer-Verlag, Berlin, Heidelberg, V. 6949, pp. 656-660.

Masip,L., Martinie,C., Winckler,M., Palanque, P., Granollers,T., Oliva, M. 2012. **A design process for exhibiting design choices and trade-offs in** (potentially) conflicting user interface guidelines. In Proceedings of the 4th international conference on Human-Centered Software Engineering (HCSE'12), Marco Winckler, Peter Forbrig, and Regina Bernhaupt (Eds.). Springer-Verlag, Berlin, Heidelberg, 53-71.

Masip, L., Granollers, T., Oliva, M. User Experience Degree and Time Restrictions as **Financial Constraints in the evaluation methodology based on heuristics**. Accepted paper in 6th Latin American Conference on Human Computer Interaction (CLIHC 2013).

Masip, L., Oliva, M., Granollers, T. **How to apply the ISO/IEC 25010:2011 in the development process of an interactive system through a methodology based on heuristics**. Sent to Behaviour & Information Technology Journal.

Masip, L., Granollers, T., Oliva, M. **Heuristic Evaluation Optimization and Adaptation towards a Semiautomatic UX Evaluation Methodology**. Sent to CHI 2014.

A previous work of this PhD about the extraction of results is published in:

González, M., Masip, L., Granollers, T., Oliva, M. 2009**. Quantitative analysis in a heuristic evaluation experiment.** *Adv. Eng. Softw*. 40, 12, 1271-1278

The feasibility analysis is published in:

Masip, L., Jurado, F., Oliva, M., Granollers, T., Trepat, T., Lozano, C. 2013 **Viabilidad de la metodología de evaluación heurística adaptada e implementada mediante Open-HEREDEUX.** *XIV Congreso Internacional en Interacción Persona-Ordenador, dentro del Congreso Español de Informática (CEDI). Madrid (Spain) pp 143-146* ***ISBN:*** *978-84-695-8352-4*

The results are sent to:

Masip, L., Conde, J., Granollers, T.,Oliva, M. **UX Measures for an Evaluation Methodology Based on Heuristics**. Sent to Interacting with Computers Journal.

Masip, L., Oliva, M., Granollers, T. **Customization of the Common Industry Format (CIF) report for the heuristic evaluation results**. HCI international conference 2014.

Bibliography

[ABA01] Abascal, J., Nicolle, C. The application of USERfit methodology to teach usability guidelines. Tools for Working with Guidelines. J. Vanderdonckt, C. Farenc (Eds.), Springer-Verlag, London pp 209-216. 2001.

[ABA01a] Abascal, J., Nicolle, C. *Why Inclusive Design Guidelines.* In C. Nicolle, J. Abascal (Eds.). Inclusive Design Guidelines for HCI. Taylor & Francis. London. pp. 3-13. 2001.

[ABR04] Abras, C., Maloney-Krichmar, D., Preece, J. User-Centered Design. In Bainbridge, *W. Encyclopedia of Human-Computer Interaction.* 2004.

[ALB96] Alben, L. Quality of experience: Defining the criteria for effective interaction design. *Interactions*; vol III pp. 11-15. 1996.

[ALL06] Allen, M., Currie, L.M., Bakken, S., Patel, V.L., Cimino, J.J. Heuristic evaluation of paper-based Web pages: A simplified inspection usability methodology. *Journal Biomedics Informatic.* Num 4, Vol. 39. pp. 412-423. 2006.

[ARD06] Ardito, C., Lanzilotti, R., Buono, P., Piccinno, A. A tool to support usability inspection. *In Proceedings of the working conference on Advanced visual interfaces* (AVI '06). ACM, New York, NY, USA, pp.278-281. 2006.

[AVI04] Avizienis, A., Laprie, J.C., Randell, B. Landwehr, C. Basic Concepts and Taxonomy of Dependable and Secure Computing. *IEEE Trans. Dependable Secur. Comput*. num 1, vol 1, pp 11-33. 2004.

[BRO88] Brown, C. M. *Human-Computer Interface Design Guidelines.* Ablex Publishing Corp. 1988.

[CAR11] Carreras, O., Sánchez C. *Checklist para la evaluación de la usabilidad web según la metodología Sirius*. Versión 3. 2011. Blog. http://olgacarreras.blogspot.com.es/2011/07/sirius-nueva-sistema-para-la-evaluacion.html

[CHE05] Chen Sherry Y. Macredie, D. The assessment of usability of electronic shopping: A heuristic evaluation. *International Journal of Information Management.* Num 6. Vol. 25. Pp 516-532. 2005.

[CHOI86] *Choice Expert VA: Decision Support Software*. Mclean, 1986.

[CON95] Constantine L. What do users want? Engineering usability into software. *Windows Tech Journal*.1995.

[COL09] Collazos C.A., Mendez, Y., Giraldo, W., Villegas, M.L, Ruiz, A. Colaboratorio de usabilidad: un mecanismo para la evaluación colaborativa de la usabilidad. *Interacción 2009*. Armenia. pp. 163-172. 2009.

[COY07] Coyle, L., Paulo, A., Xerxes, P., Vaughn, H. Evolution of a heuristic evaluation process at Bell laboratories. *In Proceedings of the Second IASTED international Conference on Human Computer interaction*. Chamonix, France : D. Cunliffe, Ed. International Association Of Science And Technology For Development. ACTA Press, Anaheim, CA, pp. 211-215. 2007.

[CRA05] Cranor, L.F., Garfinkel, S. *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly Media, 744 pages. ISBN-10: 0596008279. 2005.

[DEA06] Dearden, A., Finlay, J. Pattern languages in HCI: A Critical Review. *Journal of Human-Computer Interaction*, Vol. 21, Issue 1, pp. 49 – 102. 2006.

[DES07] Desmet P. M. A., Hekkert, P.. Framework of product experience. *International Journal of Design*. Vol 1, pp 57-66. 2007.

[DEZ09] Elena Deza & Michel Marie Deza (2009) Encyclopedia of Distances, page 94, Springer. ISBN 978-3-642-30958-8

[DIX09] Dix A., Finlay, J., Abowd, G., Beale, R. *Human-Computer Interaction*. Prentice Hall, 2009.

[EVO04] Evolucy. *Evaluación heurística de un quiosco de autoservicio de fotografía digital: caso práctico*. 2004. http://www.evolucy.com/esp/projects/evaluacion_heuristica_kiosk.html (8-7-2009)

[FAC08] Facilusar.com. http://facilusar.blogspot.com/ (19/09/2008)

[FAR01] Farenc, C., Palanque, P., Bastien, C., Scapin, D., Winckler, M. Towards a General Guidance and Support Tool for Usability Optimization. International *Conference on Universal Access in Human-Computer – UAHCI'2001*, New Orleans, USA, 5-10 August 2001.

[FOR05] Formative Test Reports. *Journal of usability studies*. Issue 1, Vol. 1, pp. 27-45, November 2005

[GON10] González, J.L. Jugabilidad: caracterización de la experiencia del jugador en videojuegos. *Doctoral Thesis.* Universidad de Granada. (2010)

[GON08] González M.P., Lorés J., Granollers, A. Enhancing usability testing through datamining techniques: A novel approach to detecting usability problem patterns for a context of use. *Information and Software Technology*, Vol. 50, pp. 547-568, 2008.

[GON06] González, M.P., Lorés, J., Pascual, A., Granollers, T. Evaluación heurística de sitios web academicos Latinoamericanos dentro de la iniciativa UsabAIPO. *Interacción'96.* pp. 143-153. 2006.

[GON09] González, J.L., Padilla, N., Gutiérrez, F. From Usability to Playability: Introduction to Player-Centred Video Game Development Process. *In Proceedings of the 1st International Conference on Human Centered Design*. Held as Part of HCI International 2009 (HCD 09), Masaaki Kurosu (Ed.). Springer-Verlag, Berlin, Heidelberg, pp 65-74. 2009.

[GON09a] González, M., Masip, L., Granollers, T., Oliva, M. Quantitative analysis in a heuristic evaluation experiment. *Advances in Engineering Software.* num 12 : Vol. 40. pp. 1271-1278. 2009.

[GRA00] Grammenos, D., Akoumianakis, D., Stephanidis, C. Integrated support for working with guidelines: the Sherlock guideline management system. Interacting with Computers. Heraklion, Greece. 3 Vol. 12. pp. 281-311.2000.

[GRA03] Granollers, T. User Centred Design Process Model. Integration of Usability Engineering and Software Engineering. Presentación Comunicación, *INTERACT 2003* (Doctoral Consortium). Zurich, SUIZA. 2003.

[GRA05] Granollers T., Lorés J., Cañas J. *Diseño de sistemas interactivos centrados en el usuario*. Barcelona : Editorial UOC, 2005. ISBN: 84-9788-320-9.

[GUT07] Gutiérrez, M. ¡Larga vida de los cajeros!". *Revista Faz.* November 2007 número 1. http://www.revistafaz.org/numero1/cajeros.pdf (19/09/2008)

[HAS06] Hassenzahl, M., Tractinsky, N. User Experience - a research agenda. *Behavior & Information Technology*, 25(2), pp 91-97. 2006.

[HEN00] Henninger, H. A methodology and tools for applying context specific usability guidelines to interface design. Interacting with Computers. Elsevier 3: Vol. 12. pp. 225-243.2000.

[INS97] Instone, K. Usability engineering for the Web. *World Wide Web Journal*. 1997. http://instone.org/node/12

[ISO25010] *ISO/IEC 25010*. Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE) System and software quality models. 2011.

[ISO25062] *ISO/IEC 25062:2006* Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Common Industry Format (CIF) for usability test reports. 2006.

[ISO9126] *ISO/IEC 9126*: Software Engineering - Product Quality.

[ISO9241] *ISO 9241-210:2010.* Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). International Standardization Organization

(ISO). Switzerland.2010.

[JEF91] Jeffries, M. R., Wharton, C., Uyeda, K. M. User interface evaluation in the real world: A comparison of four technique. *Proc. of CHI* , 119-124.1991.

[JIA09] Jiang, O., De Bruijn, O., De Angeli, A. The Perception of Cultural Differences in Online Selfpresentation. T. Gross et al. (Eds.): *INTERACT* 2009, Part I, LNCS 5726, pp. 672–685.2009.

[JOH89] Johnson, G.I., Clegg, C.W., Ravden, S.J. Towards a practical method for user interface evaluation. *Applied Ergonomics*, 20(4), 255-260. 1989.

[KEM06] Kemp, E., Setungamudalige, D.T. A resource support toolkit (R-IDE): supporting the DECIDE framework. *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's international Conference on Computer-Human interaction: Design Centered HCI*. Christchurch, New Zealand. ACM. New York. Vol. 158. pp. 61-66.2006.

[KIL09] Kılıç Delice E., Güngör Z. The usability analysis with heuristic evaluation and analytic hierarchy process. *International Journal of Industrial Ergonomics.* Num 6. Vol. 39. pp. 934-939. 2009.

[KRU05] Krug, S. *Don'T Make Me Think: A Common Sense Approach to the Web (2nd Edition)*. New Riders Publishing, Thousand Oaks, CA, USA.2005

[LAC06] Lacaze, X., Palanque, P., Barboni, E., Bastide, R., Navarre, D., From DREAM to Reality: Specificities of Interactive Systems Development With Respect to Rationale Management. In Dutoit, A.H., McCall, R., Mistrík, I., Paech, B. (eds.) *Rationale Management in Software Engineering: Concepts and Techniques*, *Rationale Management in Software Engineering*. Springer Verlag, pp.155-170. 2006.

[LAN67] Lance, G. N., Williams, W. T. A General Theory of Classificatory Sorting Strategies. 1. *Hierarchical Systems, Computer Journal*, 9, 373–380.1967.

[LAW04] Law, E.L.-C., Hvannberg, E.T. Analysis of strategies for improving and estimating the effectiveness of heuristic evaluation. *In NordiCHI 2004*, Tampere, Finland, pp. 241 – 250. 2004.

[LAW09] Law-Chong, E., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J. Understanding, scoping and defining user experience: a survey approach. *In Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*. ACM, New York, NY, USA, pp 719-728. 2009.

[LEI08] Leitner, G., Hitz, M., Melcher. R. The Role of Usability in the Design and Evaluation of Dynamic Traffic Displays. *In Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work* (USAB '08), Andreas Holzinger (Ed.). Springer-Verlag, Berlin, Heidelberg, pp 205-220. 2008.

[LER08] Lera, E., Garreta, M. Diez heurísticos emocionales. *Revista Faz*. num 2, pp. 68-71 Revista de Diseño de Interacción. Barcelona, España. 2008.

[MAG97] Maguire, M.C.. A Review of User-Interface Design Guidelines for Public Information Kiosk Systems. Stephanidis, C; Carbonell, N. (editors). *Electronic Proceedings of the 3rd ERCIM Workshop on "User Interfaces for All"*. 1997. http://ui4all.ics.forth.gr/UI4ALL-97/maguire.pdf (15/09/2008)

[MAR99] Mariage, M., Vanderdonckt, J., Pribeanu: State of the Art of Web Usability Guidelines. *In: The Handbook of Human Factors in Web Design*, pp. 688-700. Lawrence Erlbaum Associates, doi:10.1.1.58.4494, ISBN: 080584612 1999

[MAR87] Marshall, C., Nelson, C., Gardiner, M.M. Design guidelines. In: Gardiner, M.M. and Christie, B., Editors, 1987. *Applying Cognitive Psychology to User-Interface Design*, Wiley, Chichester and New York, pp. 221–278. 1987.

[MAR02] Martin, D., Rouncefield, M., Sommerville, I. Applying Patterns of Cooperative Interaction to works (Re)Design: E-Government and Planning. *In Proceedings of ACM CHI 2002*, April 20-25, 2002, Minneapolis, USA. Pages 235-242. 2002.

[MAR07] Martínez, P., Vicedo, J.L., Saquete, E., Tomás, D. *Sistemas de pregunta-respuesta Universidad de Alicante.* Departamento de Lenguajes y Sistemas Informáticos. 2007.

[MAR10] Martinie, C., Palanque, P., Winckler, M., Conversy, S. DREAMER: a Design Rationale Environment for Argumentation, Modeling and Engineering Requirements. *In proceedings of the 28th ACM International Conference on Design of Communication* (SIGDOC'2010), September 26-29, 2010, São Carlos, Brazil. ACM Press. pp. 73-80. 2010.

[MAS10] Masip, L. Análisis de viabilidad de soluciones para la automatización de la evaluación heurística. Final master Project. Univeristy of Lleida. 2010. http://www.recercat.net/handle/2072/202435

[MAS10a] Masip, L., Oliva, M., Granollers, T. Hacia la semiautomatización de la evaluación heurística: Primer paso, categorización de heurísticas, *Interacción 2010*, Grupo Editorial Garceta, Valencia, 7 to 10 of September, pp. 359-368. 2010.

[MAY92] Mayhew D.J. *Principles and Guidelines in Software User Interface Design*. New York : Prentice Hall, Inc, 1992.

[MOL90] Molich R., Nielsen J. Improving a human-computer dialogue. *Commun. ACM* . num 3, Vol. 33. - pp. 338-348. 1990.

[MOR05] Morville, P. Experience design unplugged. *In ACM SIGGRAPH 2005 Web program* (SIGGRAPH '05), Nishant Kothary (Ed.). ACM, New York, USA, Article 10. 2005.

[MUR07] Mur, C. *Diseño e implementación de un gestor para realizar evaluaciones heurísticas*. Departamento de Informática e Ingenieria Industrial ; Universitat de Lleida.2007.

[NAT07] National Institute of Standards and Technology. *Voluntary Voting System Guidelines Recommendations to the Election Assistance Commission.* Modified CIF template for voting system test laboratories-. August 2007.

[NEI07] Neil, J. Qualitative versus quantitative research: Key points in a classic debate. Research methods. 2007. On line(last visit: 19th December 2012): http://wilderdom.com/research/QualitativeVersusQuantitativeResearch.html

[NIE93] Nielsen, J. *Usability Engineering*, Published by Morgan Kaufmann, San Francisco; ISBN 0-12-518406-9 p. 26. 1993

[NIE94] Nielsen, J. Enhancing the explanatory power of usability heuristics. ACM. - New York : B. Adelson, S. Dumais, and J. Olson, *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating interdependence.* pp. 152-158. 1994.

[NIE94a] Nielsen J. Mack R.L. *Usability Inspection Methods*. New York. John Wiley & Sons. ISBN: 0-471-01877-5. 1994.

[NIE90] Nielsen J., Molich R. Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI conference on human factors in computing systems*. ed. Press ACM. - New York (USA) : [s.n.] pp. 249–56. 1990.

[NIE10] Nielsen, J., Pernice, K. *Eyetracking web usability*. Berkeley, Calif.: New Riders, 2010.

[NIE02] Nielsen J., Thair M. *Usabilidad de páginas de inicio: Analisis de 50 sitios web*. Madrid: Prentice Hall, 2002. - ISBN: 9788420532028 .

[NIE02a] Nieto, M.. La usabilidad de los terminales táctiles. *Evolucy*. 2002. http://www.evolucy.com/esp/columns/20020515_usabilidad_tactil.html (15-7-2009)

[NNG11] Website NNgroup: http://www.nngroup.com/about/userexperience.html (accessed December 2011)

[NOR88] Norman D. *The Design of Everyday Things*. New York : Basic Books, 1988.

[OLI10] Oliva, M., Masip, L., Granollers, T. Evaluación de usabilidad y accesibilidad de un conjunto de dispositivos interactivos denominados Puntos de Información Ciudadana. S*cire: representación y organización del conocimiento* ISSN: 1135-3716. 2010.

[PAL07] Palanque P., Lacaze X. DREAM-TEAM: A Tool and a Notation Supporting Exploration of Options and Traceability of Choices for Safety Critical Interactive Systems. *In Proceedings of INTERACT 2007*, LNCS, Springer Verlag, p. 230-245. 2007.

[PEA95] Pearson, K. *Mathematical Contributions to the Theory of Evolution.* - III. Regression, Heredity, and Panmixia. University College, London. Communicated by Professor HENRICI, F.R.S.1895.

[PIE04] Pierotti D. *Heuristic Evaluation - A System Checklist.* Xerox Corporation. 2004. http://www.stcsig.org/usability/topics/articles/he-checklist.html

[PIR11] Pirker. M. Enhancing and Evaluating the User Experience of Interactive TV Systems and their Interaction Techniques. *EuroITV2011*, June 29th-July Lisbon, PORTUGAL. 2011.

[PON06] Ponsa, P., Díaz, M. Creación de guía ergonómica para el diseño de interfaz de supervisión. VII *Congreso Internacional de Interacción Persona-Ordenador*. Puertollano, Ciudad Real.2006.

[PRA00] Prates, R., de Souza, C., Simone, D.J. Methods and tools: a method for evaluating the communicability of user interfaces. *Interactions* 7, 1, 31-38. 2000.

[PRE94] Preece, J., Rogers, Y. *Human-Computer Interaction*, Addison-Wesley.1994.

[ROD02] Rodrigo, L., García, A., Martínez, P. Planeamiento semántico y pragmático para gestión de diálogos en asistentes virtuales. *Procesamiento del Lenguaje Natural*, Revista, num.28. 2002.

[ROT11] Roto, V., Vermeeren, A., Väänänen-Vainio-Mattila, K., Law. E. User experience evaluation - which method to choose?. *In Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part IV* (INTERACT'11), Pedro Campos, Nuno Nunes, Nicholas Graham, Joaquim Jorge, and Philippe Palanque (Eds.), Vol. Part IV. Springer-Verlag, Berlin, Heidelberg, 714-715. 2011.

[SAA80] Saaty T.L. *The Analytic Hierarchy Process.* McGraw-Hill International Book, 1980.

[SAU12] Sauro, J. *The Value Of Multiple Evaluators In Heuristic Evaluations.* http://www.measuringusability.com/blog/he-multiple.php

[SAU12a] Sauro, J. *How Effective Are Heuristic Evaluations?* http://www.measuringusability.com/blog/effective-he.php

[SCH10] Schumacher, M., Lowry, S. *Customized Common Industry Format Template for Electronic Health Record Usability Testing.* National Institute of Standards and Technology Interagency/Internal Report (NISTIR) – 7742 November 16. 2010.

[SHN87] Shneiderman B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison-Wesley Publ. Co. 1987. ISBN:0201165058.

[SID08] Fundación Sidar. *Principios del Diseño Universal o Diseño para Todos*. http://www.sidar.org/recur/desdi/usable/dudt.php (19/09/2008)

[SIT03] Sitepal. *Speaking avatar best practices guide*. 2003. Allowed in http://www.sitepal.com/pdf/bestpractice_2008.pdf

[SMI86] Smith S., Mosier J. *Guidelines for designing user interface software*. Bedford : Mitre corp. 1986.

[SOL12] Solano, A.*Propuesta metodológica para la evaluación colaborativa de la usabilidad de aplicaciones de Televisión Digital Interactiv.* Tesis de Maestría, Instituto de Posgrados en Ingeniería Electrónica, Universidad del Cauca, Popayán, Cauca, 2012.

[SPI12] Spinhof, L. *UserPlus Advisor* http://www.userplus.org, created by

[SQU09] Squac Grupo Accusa. Herramienta de soporte para las evaluaciones herurísticas. Last visit: 02-12-2010. http://squac.iti.upv.es/blog/2009/accusa-una-herramienta-de-soporte-a-la-ingenieria-de-usabilidad-y-accesibilidad/.

[STE99] Stephanidis, C. Akoumianakis, S., Accessibility guidelines and scope of formative HCI design input: contrasting two perspectives, in *Proc. of the 5th ERCIM Workshop on User Interfaces for All* (Dagstuhl, 28 novembre-1 décembre 1999), GMD Report 74, GMD - Forschungszentrum

Informationstechnik GmbH, Sankt Augustin, 1999, pp. 119-130. 1999.

[SUS09] Sussex Universidad de Interactive Heuristic Evaluation. Interactive Heuristic Evaluation. 02-11-2009. http://www.id-book.com/catherb/.

[SWA07] Sward, D., MacArthur, G. Making user experience a business strategy. In E. Law et al. (eds.), *Proceedings of the Workshop on Towards a UX Manifesto*, 3 Sept. 2007, Lancaster, UK pp 35-40. 2007.

[THE99] Thevenin, D., Coutaz, J. Plasticity of User Interfaces: Framework and Research Agenda. *Proceed. of INTERACT'99. Conference on Human-Computer Interaction*, Vol. 1, pp. 110-117. 1999.

[TOG03] Tognazzini B. *First principles of interaction design. Interaction design solutions for the real world*. AskTog, 2003. http://asktog.com/basics/firstPrinciples.html.

[TRA09] Travis, T. Expert Review Checkpoints", User Focus. 2009. http://www.userfocus.co.uk/resources/guidelines.html

[UNE139801] UNE139801 (2003). Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad al ordenador. Hardware. AENOR. Depósito legal: M 42219:2003.

[UNE139802] UNE139802 (2003). Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad al ordenador. Software. AENOR. Depósito legal: M 42220:2003

[UPA09] "UPA 2009 Salary Survey" Usability Professional's Association.

http://usabilityprofessionals.org/usability_resources/surveys/2009salarysurvey_PUBLIC.pdf

[USA08] Usable y Accesible. http://olgacarreras.blogspot.com/ (19/09/2008)

[USD06] US Department of health and human services. Research-based web design and usability guidelines. http://web.archive.org/web/20060827054938/http://www.usability.gov/pdfs/chap.html

[UXN11] Website UXnet: http://www.uxnet.org/ (accessed December 2011)

[VAN95] Vanderdonckt, J. Accessing guidelines information with SIERRA. *In Proceedings of Fifth IFIP TC 13 International Conference on Human‑Computer Interaction*,Interact'95. Ed. K. Nordbyn P.H. Helmersen, D.J. Gilmore and S.A. Arnesen. London. Chapman & Hall. pp. 311. 1995.

[VAN99] Vanderdonckt, J. Development milestones towards a tool for working with guidelines. *Interacting with Computers* 12(2): pp. 81-118. 1999.

[VAN99a] Vanderdonckt, J. Automated Generation of an On-Line Guidelines Repository. HCI (1) pp 998-1002. 1999.

[VAN01] Vanderdonckt, J., Farenc, C. (Eds.) *Tools for Working with Guidelines*. Springer-Verlag, London. 2001.

[VER10] Vermeeren, A., Lai-Chong Law, E, Roto, V., Obrist,M, Hoonhout,J. &Väänänen-Vainio-Mattila, K.User Experience Evaluation Methods: Current State and Development Needs. *In Proc. of NordiCHI'10*, ACM. 2010.

[VOG01] Vogt, T. *Difficulties in Using Style Guides for Designing User Interfaces*. Tools for Working with Guidelines. J. Vanderdonckt, C. Farenc (Eds.), Springer-Verlag, London, pp. 197-208. 2001.

[VON04] von Baggo, K., Johnston, L., Burmeister, O., Bentley, T. Common Industry Format: Meeting Educational Objectives and Student Needs? *In proceeding of: Computer Human Interaction*, 6th Asia Pacific Conference, APCHI, LNCS 3101, M. Masoodian et al. (Eds.): Rotorua, New Zealand, pp. 521-530, June 29 - July 2, 2004.

[WAR63] Ward, J. H., Jr. Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58, 236–244. 1963.

[W3C08] W3C (2008). Web Content Accessibility Guidelines 2.0. W3C Candidate Recommendation April 2008.[http://www.w3.org/TR/WCAG20/]

The author of this thesis is supported by a pre-doctoral fellowship from the University of Lleida.

This thesis has received a grant for linguistic revision from the Linguistic Service at the University of Lleida (2013 call).

Appendix A

Specification of the usability categories (in Spanish)

Categoría	Autor	Nomenclatura utilizada por el autor
Consistencia	Nielsen/Pierotti	Consistencia y estándares
	Marshall et al./ Instone/ Tognazzini	Consistencia
	Shneiderman	Esforzarse por la consistencia
	Molich et al.	Hacerlo consistente
	Constantine	Estructura/Reutilización
	Mayhew	Consistencia y robustez/ Compatibilidad del usuario, del producto, de las tareas y de los procesos del sistema
	Tognazzini	Objetos humanos
Feedback	Nielsen/Pierotti	Visibilidad del estado del sistema
	Smith et al.	Orientación del usuario: Información de estado/Realimentación rutinaria/Realimentación del error/ Transmisión de datos: Control de la transmisión
	Shneiderman	Ofrecer retroalimentación/Diseñar el diálogo para mostrar el trabajo pendiente
	Molich et al.	Proporcionar retroalimentación
	Constantine/ Instone/Marshall et al.	Retroalimentación
	Mayhew	Sensibilidad y retroalimentación
	Tognazzini	Autonomía
Gestión de errores	Nielsen/Instone/ Pierotti	Prevención de errores
	Tahir et al.	Comunicación de problemas técnicos y gestión de emergencias
	Pierotti	Ayudar a los usuarios a reconocer, diagnosticar, y recuperarse de un error
	Smith et al.	Control de secuencias: Gestión de errores/Alarmas/ Entrada de datos: Validación de datos
	Marshall et al.	Procesamiento de errores
	Brown	Mensajes de error y asistencia en línea
	Shneiderman	Ofrecer una gestión de errores simple
	Norman	Diseñar para los errores/ Cuando todo falla, estandarizar
	Molich et al.	Proporcionar unos buenos mensajes de error/Prevención de errores
	Constantine	Tolerancia
	Instone	Mensajes de error correctos
	Tognazzini	Proteger el trabajo/ Guardar el estado
Reducir la carga de memoria	Nielsen/Pierotti	Reconocimiento más que recuerdo
	Constantine	Visibilidad
	Shneiderman	Reducir la carga de memoria a corto plazo

Categoría	Autor	Nomenclatura utilizada por el autor
	Molich et al./ Instone	Minimizar la carga de memoria del usuario
	Tognazzini	Uso de metáforas/Anticipación
	Norman	Hacer las cosas visibles/ Conseguir los mapas acertados/ Explorar el poder de las limitaciones
Flexibilidad	Nielsen	Flexibilidad y eficiencia de uso
	Pierotti	Flexibilidad, estética y diseño minimalista/ Interacción respetuosa y gratificante con el usuario
	Marshall et al.	Adaptación
	Mayhew	Flexibilidad
	Tognazzini	Daltonismo/Legibilidad
	Constantine	Soporte
Diálogo	Nielsen/Pierotti	Unión entre el sistema y el mundo real
	Tahir et al.	Redacción de contenido/La URL
	Marshall et al.	Seleccionar términos, palabras y objetos:
	Brown	Redacción efectiva/ Diseño de diálogos
	Molich et al.	Diálogo simple y natural
	Molich et al./ Instone	Hablar el lenguaje del usuario
	Norman	Utiliza el conocimiento del mundo y el de la cabeza
	Tognazzini	Eficacia del usuario
Control del usuario	Nielsen/Pierotti	Control y libertad para el usuario
	Smith et al.	Control de secuencias: Selección de transacciones/Interrupción/ Transmisión de datos: Iniciar la transmisión
	Marshall et al.	Locus de control
	Brown	Control y visualización de recursos
	Shneiderman	Soportar el control por el usuario/ Permitir deshacer fácilmente las acciones realizadas
	Constantine	Progresión
	Mayhew	Manipulación directa/ Control/ WYSIWYG
Facilidad de uso	Pierotti	Habilidades
	Marshall et al.	Diseño de procedimientos y tareas/ Analogía y metáforas/ Entrenamiento y práctica/ Unión entre usuario y tarea
	Norman	Simplificar la estructura de las tareas
	Constantine/ Mayhew	Simplicidad
	Mayhew	Familiaridad/ Facilidad de uso y aprendizaje
	Tognazzini	Aprendizaje
	Constantine	Acceso
Atajos	Tahir et al.	Herramientas y accesos directos a tareas
	Shneiderman	Proporcionar atajos para los usuarios frecuentes
	Molich et al.	Proporcionar atajos
	Constantine	Eficacia
	Instone	Atajos
Ayuda	Pierotti/Instone	Ayuda y documentación
	Smith et al.	Orientación del usuario: Ayudas de trabajo/ Cambiar el diseño
Navegación	González et al.	Navegación: Áreas de Navegación/Orientación
	Marshall et al./ Tahir et al.	Navegación

Categoría	Autor	Nomenclatura utilizada por el autor
	Tognazzini	Navegación visible/ Interfaces explorables
Protección	Mayhew	Protección
	Pierotti	Privacidad
	Tahir et al.	Recopilación de datos del cliente
	Smith et al.	Protección de datos: General/ Identificación del usuario/ Acceso a datos/ Entrada/Modificación de datos/ Transmisión de datos/ Cambiar el diseño
Salidas de emergencia	Molich et al.	Proporcionar marcas claras
	Instone	Marcar claramente las opciones de salida
Búsqueda	González et al.	Búsqueda: Área de búsqueda/ Resultado de la búsqueda
	Tahir et al.	Búsqueda
Internacionalización	González et al.	Contenido: internacionalización
	Tahir et al.	Fechas y horas en formato internacional
Contenido	González et al.	Contenido: Información/Imágenes
	Marshall et al.	Diseño de la pantalla / Organización / Interacción multimodal y multimedia
	Smith et al.	Visualización de datos: Texto/ Formulario de datos/ Tablas/ Gráficos/ Formato/ Codificación/ Control de la pantalla/ Cambio de diseño

Appendix B

Example of communicability heuristics

These heuristics are for designers. So, they are written as declarative sentences.

Dialog

- The system should use affirmative sentences
- Sentences should be without any ambiguity
- Vocabulary should be familiar to the users
- Terms for a jargon should not appear
- Information should be presented in a consistent way with consistent structures and grammatically correct
- Information should be presented in a logical way
- Uppercase mode should be used only when it is essential
- Punctuation should be used correctly, avoiding exclamation marks
- The use of prefixes that qualify the meaning of the sentence should be avoided
- The dialogs with the main information should be shown as simplified as possible
- Ending lines with broken words should be avoided
- If it is essential to remember an item, it should be located in a highlighted and unusual position
- The entire text should be presented in the same page to avoid broken text in more than one page
- Sentences should be short and simple
- The main ideas should be in the first lines of the text
- Sentences should be in the active voice
- Repeting the content should be avoided
- If the content is clear enough, the part of content should not be labeled
- Lists and categories with one option should be avoided
- An imperative mode for required actions should be used
- The system should have an understandable URL

Abbreviations

- Abbreviations should be used only if they have a minor meaning
- A different abbreviation should be used for each word
- The way of doing the abbreviation should always be the same
- A dictionary of abbreviations should be provided
- The abbreviations should only be used with the needed information

- Abbreviations that are not very clear should be avoided
- The meaning of the abbreviations should be immediately presented after the first occurrence
- Abbreviations should not include punctuation

Conversation

- The system should enable inserting questions in different languages
- The system should have audio
- The system should provide information about itself
- The system messages should be in rotation, always different or constantly changing
- Sentences should be coherent and they should fit the questions
- If the system does not have the needed information, the system should inform the user about it
- User should be able to contextualize the question and the system should present the same results
- The system should be able to relate different questions to delimit the results of the search
- The system should identify a change of the topic

Example of recoverability heuristics

Protection

- In multiple data entries, making a partial storage should be possible
- Data should be stored when users need help
- It should be possible to do an update and not another insert after an error

Data transmission

- A copy of the transmitted message should be saved

Appendix C

CIF Template

Common Industry Format for Usability Test Report v1.1

Comments and questions about this format: iusr@nist.gov

[❖ Name the product and version that was tested]

[❖ Who led the test]

[❖ When the test was conducted]

[❖ Date the report was prepared]

[❖ Who prepared the report]

For:

[❖ Customer Company Name]

[❖ Customer Company contact person]

Address inquiries to: [❖ Contact name(s) for questions and/or clarifications]

Phone: [❖ Enter phone number]

Email: [❖ Enter email address]

Address: [❖ Enter mailing or postal address]

Executive Summary

This report provides information on the user experience test carried out through an inspection method based on heuristics in the NAMEINTERACTIVESYSTEM.

Your interactive system, the NAMEINTERACTIVESYSTEM, is presented as DESCRIPTIONINTERACTIVESYSTEM .

This analysis was carried out using Open-HEREDEUX (Open HEuristic Resource for Designing and Evaluating User eXperience). It enables us to achieve the best heuristics for a specific interactive system such as TYPESYSTEM, it supports the evaluation of this set of heuristics by different evaluators (including the end or real user of the system) and finally, Open-HEREDEUX creates this document automatically using the information inserted during the whole process.

The evaluation of your interactive system was done on DATEOFEVALUATION by RESPONSIBLEGROUP.

The methodology to evaluate your interactive system was carried out by NUMBEREVALUATORS evaluators. Each one individually interacts with the system and checks the set of needed heuristics according to the following goals:

GOALS

The evaluation results present positive and negative aspects to be considered in the interactive system. The main positive aspects are:

POSITIVEASPECTS

And the main aspects to be improved are:

NEGATIVEASPECTS

In summary, the NAMEINTERACTIVESYSTEM presents AMOUNTU of problems in the U degree, UUAMOUNT of problems in the UU degree and XAMOUNT of problems in UUU degree.

Introduction

Full Product Description

[❖ Formal product name and release or version]

[❖ Describe what parts of the product were evaluated]

[❖ The user population for which the product is intended]

[◆ Any groups with special needs]

[◆ brief description of the environment in which it should be used]

[◆ the type of user work that is supported by the product]

Test Objectives

Bearing in mind that user experience considers all aspects, internal as well as external of the user and interactive systems, which provoke any feeling in whoever uses the interactive system in a specific context of use, the main goal of this test is that end users get a positive experience as possible.

Specifically, the goals of this test are:

GOALS

Method

Participants

The test was carried out for NUMBEREVALUATORS. The following tables show some features of the evaluators but personal information is omitted to preserve the privacy of the evaluators:

On the one hand, there are HCIEXPERTS HCI experts. HCI experts shall get results for the other UX facets selected for this evaluation. The specific results are presented in following sections.

	Gender	Age	Education	Computer experience	Years in HCI	Amount of evaluations done	Product experience
1	GENDER 0	EDAD 0	EDUCATION 0	COMPUTER 0	HCI0	AMOUNT 0	ISEXPERIENCE 0
9	GENDER 8	EDAD 8	EDUCATION 8	COMPUTER 8	HCI8	AMOUNT 8	ISEXPERIENCE 8
10	GENDER 9	EDAD 9	EDUCATION 9	COMPUTER 9	HCI9	AMOUNT 9	ISEXPERIENCE 9

On the other hand, there are ENDUSERS end users who are experts in the system but they are not experts in HCI aspects. Therefore, this type of users shall present the results of the emotional facet.

	Gender	Age	Education	Computer experience	Product experience
1					
2					
..					
n					

Context of Product Use in the Test

Test Facility

Every participant or evaluator carried out the evaluation individually. Each evaluator connects to Open-HEREDEUX (it is detailed in the two sections below) using his/her own computer/laptop/tablet. There are not strict requirements for doing the evaluation under special environment conditions.

The only requirement is that the participant should try to execute the evaluation in a quiet place and without external interruptions to obtain the most objective results.

Participant's Computing Environment

[❖ Computer configuration, including model, OS version, required libraries or settings]

[❖ If used, browser name and version; relevant plug-in names and versions]

	Operative System	User uses the technology through	Navigator
1	OPSYS0	DEVICE0	AVN0
2	OPSYS1	DEVICE1	AVN1
3	OPSYS2	DEVICE2	AVN2
4	OPSYS3	DEVICE3	AVN3
5	OPSYS4	DEVICE4	AVN4
6	OPSYS6	DEVICE5	AVN5
7	OPSYS6	DEVICE6	AVN6
8	OPSYS7	DEVICE7	AVN7
9	OPSYS8	DEVICE8	AVN8
10	OPSYS9	DEVICE9	AVN9

Display Devices

[❖ If screen-based, screen size, resolution, and color setting]

[❖ If print-based, the media size and print resolution]

El texto presentado no puede ocupar más de una línea

Audio Devices

El texto presentado no puede ocupar más de una línea

Manual Input Devices

[⬥ If used, specify the make and model of devices used in the test]

Test Administrator Tools

The tool used to carry out the user experience test is called Open-HEREDEUX (Open HEuristic REsource for Designing and Evaluation User eXperience). Open-HEREDEUX is composed of four components:

(i) The Open Repository that saves all the needed information for (ii) the Adviser of heuristics who gets the list of heuristics as adequate as possible for designers and for evaluators. Then, if an evaluation of the UX would be carried out, (iii) the Scorer of heuristics can be used. Finally, the Scorer of heuristics saved all the information to send it to (iv) the Results Analyzer who reaches quantitative and qualitative results.

The whole set of heuristics is attached in Appendix A.

Experimental Design

The methodology used to carry out the UX test is based on the heuristic evaluation methodology. The heuristic evaluation as an inspection methodology presents three main steps:

- The preparation of the evaluation where the head of the project should select the best heuristics for the specific interactive system, the evaluators and the severity factors.
- The scoring process where the evaluators can mark individually the whole set of heuristics using the preselected severity factors.
- The results extraction where qualitative and quantitative results are presented.

There are three variables to reach in the test: the correlation among evaluators, the user experience degree and the percentage of user experience. This information is obtained according to the scores of each participant in each heuristics. These scores are called severity factors and are defined through the impact and the frequency.
In the “Usability metrics” section the definition of these variables are detailed.

Procedure

The responsible of the project send an email to the participants giving the needed instructions to provide the participants with the essential information to carry out the test without problems.

When each participant has enough time to carry out the test, each one individually connects to the Open-HEREDEUX interface to start the test.

The first step of the test is to fill in a questionnaire to check the participant profile. Then, participants are invited to use the interactive system for some time (5-10

minutes) without answering the heuristics. After this time, they are more familiar with the system and they can start the scoring of heuristics.

The methodology based on heuristics gives freedom to participants. So, they can do what they like in the system to answer the whole set of heuristics.

Participant General Instructions

Please see Appendix B of this document for the template of the email that the responsible of the evaluation send to the participants.

Please see Appendix C for the participant profile questionnaire.

UX Metrics

Correlation among Evaluators (CaE)

The Correlation among Evaluators (CaE) is defined to know if participants score more or less equal each heuristic.

User eXperience Degree (UXD)

The User eXperience Degree (UXD) is the level of importance that a heuristic has in a specific interactive system. So, following the degrees considered in the accessibility, three UXDs are proposed:

- **U degree**: the heuristics of the U degree are essential to assure that the user who will use the interactive system will get a positive experience.
- **UU degree**: the heuristics of the UU degree are necessary to assure that the user who will use the interactive system will get a positive experience.
- **UUU degree**: it is advisable to consider the heuristics of the UUU degree to assure that the user who will use the interactive system will get a positive experience.

Percentage of User eXperience (PUX)

The Percentage of User eXperience (PUX) is the amount of heuristics that are not a problem taking as a reference the total amount of heuristics.

Results

Data Analysis

Severity factors

The evaluators use two marks to score each heuristics. Generally, these marks are called severity factors. Using both severity factors is possible to obtain the metrics presented above:

- **Impact:** The impact of the problem if it occurs: Will it be easy or difficult for the users to overcome?
  - 0→I don't agree that this is a problem at all
  - 1→Cosmetic problem only: needs to be fixed unless extra time is available on project
  - 2→Minor problem: fixing this should be given low priority
  - 3→Major problem: important to fix, so it should be given high priority
  - 4→Catastrophe: imperative to fix this before product can be released
- **Frequency:** The frequency with which the problem occurs: Is it common or rare?
  - 0→Never
  - 1→Rarely
  - 2→Sometimes
  - 3→Usually
  - 4→Always

Correlation among Evaluators (CaE)

As we commented on before, the test was carried out by NUMBEREVALUATORS .

The index of correlation among them is the following ones:

INCOR

The next graph shows the Pearson Correlation Graphic:

GRAPHCORRELATION

Types of problems

According to the following graphs, the whole set of heuristics and its punctuations are divided into 4 different types of problems:

- There is not a problem.
- It is a minor problem.
- It is an important problem.
- It is a catastrophe.

The following graph shows the hierarchical classification of the problems:

GRAPHCLUSTERING

TYPEGROUPS

Presentation of the Results

In this section, the results are presented in two ways. First of all, qualitative results are detailed through a list of improvements to apply in the interactive system. Then, qualitative results are presented regarding the UXD and PUX.

Qualitative

LISTIMPROVEMENTS

Quantitative

As was presented in the previous section, the Percentage of User eXperience (PUX) is the amount of heuristics that are not a problem taking as a reference the total amount of heuristics.

According to this definition in the following lines the percentage of UX in each UX degree is presented.

The Percentage of UX in the U degree is: UPERCENTAGE %.

The Percentage of UX in the UU degree is: PERCENTAGEUU %.

Finally, the Percentage of UX in the UUU degree is: PARTIALUUU %.

In summary, the total Percentage of UX of the system is: TOTALUX %.

Appendices

Appendix A

Entire list of heuristics used in this evaluation:

WHOLELISTOFHEURISTICS

Appendix B

Email sent to the participants

Dear colleague,

It is our pleasure to inform you that you are one of the selected participants to evaluate our interactive system according to the project PROJECT.

We have selected you as a participant because we consider that your experience will be very useful for our study. It is important to highlight that we do not want to evaluate your personal skills using the system or the way that you will use it. We would like to consider your answers to improve the interactive system according to your marks and comments.

Please, connect to Open-HEREDEUX to start the evaluation and follow the instructions presented there. The basic steps to carry out the study are:

1. Connect to the Scorer of the Open-HEREDEUX.
2. Select the project: PROJECT.
3. Fill in the questionnaire to know you profile
4. Use the interactive system for a few minutes (5-10 minutes) to be familiar with it.
5. Score the set of heuristics.

If you have any doubt, comment, question or something else, do not hesitate to contact me.

Thank you very much for your time and participation.

Appendix C

Participants profile questionnaire

Gender: □ Male □Female

Age: □ 18-25 □ 26-35 □36-45 □46-55 □56-65 □More than 65

Education:

□ High School □ University Degree □Master Degree □Doctorate

How often do you use computers?

□Once a week □Twice a week □ Three-four times □ Everyday

Experience in Human-Computer Interaction:

□ 0 years □ 1-5 years □ 6-10 years □ More than 10 years

Experience in the interactive systems:

□ 0 years □ 1-5 years □ 6-10 years □ More than 10 years

Amount of evaluations done:

□ 0 evaluations □ 1-5 evaluations □ 6-10 evaluations □ More than 10 evaluations

If the interactive system is a Website application, please answer these questions:

You use the interactive system through:

□ Computer □ Tablet □ Mobile phone

Operative System:

□ Windows □ iOs □ Linux □Others: _____

Browser:

□ Explorer □ Firefox □ Opera □ Others: _____