Visualizing Usability

Filipe Manuel Nunes Morgado

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisor: Prof. Sandra Pereira Gama

Examination Committee

Chairperson: Prof. Francisco João Duarte Cordeiro Correia dos Santos
Supervisor: Prof. Sandra Pereira Gama
Members of the Committee: Prof. Daniel Jorge Viegas Gonçalves

June 2018

Acknowledgments

I would like to thank my parents, António and Fernanda, for their friendship, encouragement and caring over all these years, for always being there for me through thick and thin, for being extremely supportive and without whom this project would not be possible. I would also like to thank my grandparents, aunts, uncles and cousins for their understanding and unconditional support throughout all these years. A specially thank you to my uncle Manuel, who passed away recently, who taught me to be persistent and to not give up on any of my goals.

I would also like to acknowledge my dissertation supervisors Prof. Sandra Gama for the insight and suggestions provided throughout this these months and for being available to help as soon as possible. Your support and sharing of knowledge and experience has made this Thesis possible.

Last but not least, to all my friends that were always there for me to cheer me up and make me laugh during hard times. And to you, Filipa, who always believed in my capabilities and accompanied me through this journey and final chapter of my student life. Thank you.

I'm sure I skipped someone, but to those who I did miss – Thank you.

Abstract

While developing a product and more specifically an interface, measuring User Experience and how users interact with it is imperative, since a poorly usable interface can affect users' willingness to interact with it and cause a decline in the overall user interest and satisfaction. The measurement and analysis of Usability related metrics can sometimes be hindered due to the nature and complexity of this data, resulting in a poor extraction of such information. Therefore, Information Visualization can be used as a mechanism to better study and detect product deficiencies at an early stage and represent a powerful tool to improve the analysis of usability and complement user testing, providing insight for changes and, ultimately, an interface with a suitable and pleasurable experience for users. This work addresses ways of exploring and correlating usability metrics and possibly identifying patterns through the usage of Visualizations, which can potentially help draw "high level" conclusions based on both derived and statistical measures, and thus contributing to better results while designing interfaces. Furthermore, the presented work includes a validation approach through both Usability Testing and Case Studies on the developed solution.

Keywords

User Experience; Information Visualization; Analysis; Usability; User Testing.

Resumo

Durante o desenvolvimento de um produto e mais concretamente uma interface, a medição da Experiência do Utilizador e de como este interage com esta é imperativa, porque uma interface com baixa usabilidade pode afectar a facilidade de interação com esta e causar um declínio no interesse dos utilizadores. A medição e análise de métricas de Usabilidade pode, por vezes, sofrer devido à natureza e complexidade associadas a este tipo de dados, resultando numa extração de informação de baixa qualidade. Assim, a Visualização de Informação pode ser usada como um mecanismo para facilitar a análise e a deteção de deficiências no produto num estado preliminar, disponibilizando uma forte ferramenta para a melhoria da análise da Usabilidade e complementar a fase de testes com utilizadores, providenciando informação sobre como melhorar a interacção com estes e, por fim, uma interface com uma experiência adequada e agradável para os utilizadores. Este trabalho pretende estudar formas de explorar e correlacionar métricas de Usabilidade e possivelmente identificar padrões através da utilização de Visualizações, tendo potencial para extrair conclusões de ”nível elevado” que seriam de outra forma difíceis de obter, baseando-se em métricas derivadas e estatísticas e portanto contribuíndo para melhores resultados no desenho de interfaces. Além disso, este trabalho ínclui um método de validação assente através de Testes com Utilizadores e Casos de Estudo sobre a solução desenvolvida.

Palavras Chave

Experiência do Utilizador; Visualização de Informação; Análise; Usabilidade; Testes com Utilizadores.

Contents

List of Figures

List of Tables

Acronyms

infoVis	Information Visualization
HCI	Human Computer Interaction
CG	Computer Graphics
POC	Proof of Concept
SUS	System Usability Scale
UCD	User Centered Design
UX	User Experience

1

Introduction

Contents

Usability is very important in a way that it helps designers and developers to better understand how their interface presents itself and interacts with the users, as well as ensuring that it is both easy to learn, effective and enjoyable to be used, providing an overall positive experience. Disregarding usability as a whole can compromise the overall interaction between an interface and users, since if it is hard to learn and to manage, they will often leave and look for other options. These situations can be managed and improved by observing customers and by involving them in the design process, allowing developers to introduce a more stable and ready final product in the market. Some of this approaches share the intent of retrieving feedback and performance metrics, in order to evaluate how easily users interact with the design, but the problem resides in the difficulty in extracting information and analyzing data with such complexity and nature.

Thus, Information Visualization represents a powerful tool to present data in an intuitive way, as well as helping users improving their decision making and drawing conclusions based on exploration and correlation of data. In this context, by providing an alternative of assessing usability data relying on Information Visualization, one can ease the process of extracting important information and identifying certain critical areas where the interface needs improvement. This is possible by appropriately studying user testing data through the analysis of both derived and statistical measures.

1.1 Motivation

Research on this topic was made and a shortage of work relating Usability and Information Visualization was found. Thus, the assessment of usability in this context represents a new challenge and hopefully a new way to explore these areas. Usability is commonly studied through the analysis and statistical work done on user testing results, which can sometimes be misleading due to the subjective nature of some metrics, and Information Visualization can further improve and ease these processes. It is known that Information Visualization is used as a way to provide contextual information, but there must be concerns regarding how it is visually encoded: as expressed by Munzner [1], some representations might incur additional cognitive workload to the user, rather than the optimal, which is reducing it. Ideally, a visualization is developed in order to reduce the cognitive load associated with the analysis of the metrics represented, and to provide users with intuitive ways to answer their questions, as well as finding patterns and correlations between the datasets. Therefore, an interactive visualization regarding this type of data can, ultimately, help identify problems during the early design stages and force developers to address concerning areas more effectively.

1.2 Objectives

The main objective of this work is to **Study interactive ways of assessing Usability measures through Information Visualization**. In order to reach this goal, a number of other tasks were achieved along with the development process, such as defining a conceptual idea for an Information Visualization Dashboard and which questions were the most important for the visualization to answer, as well as how the data used would be mapped. The design of the proposed solution and changes made throughout the several design iterations relied on a User Centered Design (UCD) approach and therefore each option and decision made during its evolution was substantiated accordingly. Furthermore, the interactive dashboard solution went through a series of formative evaluations in order to resolve any problems found by the users, culminating with Summative evaluation through Usability testing with users and three case studies made with Information Visualization experts. Statistical analysis on the results obtained were also presented in hopes of obtaining additional insight on the way users interact with the solution and to identify potential flaws on the design, as well as deriving conclusions from the results collected.

Due to mostly time constraints and the possibility of maximizing the information obtained from this study and the quality of the solution, there was a decision to sustain its development on a *Proof of concept* approach and therefore disregarding the usage of real data. Since gathering and cleaning real user data would consume extra time while not adding nor subtracting any substantial value to the main purpose of the presented thesis, this option was discarded.

1.3 Document Structure

The presented document is structured as follows. Chapter 2 includes a background study and fundamental concepts used in this thesis, and the importance of Usability to its standards, related approaches and field usages. Chapter 3 represents an investigation and study done on related visualization work divided into three different categories, as well as a comparative discussion on the qualities and deficiencies each said category possesses, providing additional information and an overall insight on the conditions and requirements the solution must comply with. Every piece of information regarding the prototype's development is written on Chapter 4, such as the approaches adopted; decisions regarding the Dataset and correspondent mapping; an high level description of the architecture; metrics and questions used for the evolution of our proposal. Along with the previously stated processes, the evolution and growth of our prototype is shown along with its justified changes used to create the final iteration. Chapter 5 describes the evaluation processes used for quantitative studies, the results and feedback collected from test participants, as well as the three Case Studies and the insight obtained. Chapter 6 is reserved for the limitations found throughout this study; what future additions and changes should be made in order to further improve the quality of the solution and conclusions regarding it.

2

Background

Contents

In this Chapter, Usability will be taken in consideration as a crucial factor that shapes User Experience and contributes to a better design solution. The meaning of usability and its most used metrics, how important it is while developing an interface and why the user should be equated in the whole process of development are all topics discussed in the following paragraphs. Furthermore, usability and user experience goals, evaluation techniques, derived approaches and other uses for usability are also studied and discussed in this Chapter.

2.1 Usability - What is it

In short, Usability is a very important component of interaction design and it is generally regarded as ensuring that the interactive interface is both easy to learn, effective and enjoyable to be used, providing an overall positive experience for users. The usability improvement is made by optimizing the interactions people have with the interface, making it easier for them to carry out their daily tasks, whether it is at school, work or in their everyday life.

To more specifically study usability, its six main goals are represented in the following list, as mentioned by Preece et. al [2]:

- Effectiveness – Effectiveness represents how good a system is at carrying out the initial proposed tasks and, therefore, producing the desired result. It refers to the completeness and accuracy at which the users achieve specified goals.

- Efficiency – Efficiency can be described as the quantity of resources used to carry out a task, such as time or clicks used to reach a goal. Poor efficiency is generally associated with an increase in time and effort for the user.

- Safety – Involves protecting the user from hazard, such as dangerous conditions or undesirable situations. It also refers to the perceived fears the users might have of the consequences of making errors while interacting with the product. High levels of safety in a system should boost the users' confidence and allow them to fully explore an interface by trying more difficult tasks or operations.

- Utility – This goal measures if the system provides the right kind of functionality so that users can perform the tasks they need in a way they want to. Therefore, a system with a low value of utility may result in the user not being able to accomplish a give task.

- Learnability – No one likes to spend a long time learning how to use a system, thus learnability is a very important factor, since experience shows that users prefer to explore the system hands on, rather than dealing with intensive manual studying, as Dix et. al notes [3]. It refers to how easy it is for the user to learn the system and perform tasks without too much effort.

- Memorability – The last goal, memorability represents how easy a system is to use, once learned. It is especially important for interactive systems that are not used frequently. Thus, if a user, for instance, has not used the system in a couple of months, he should be capable of resuming the normal usage of the system without much effort.

It is important to note that Usability itself can be described by a different number of goals. Taking for example the International Organization for Standardization 9241-11 [4], Usability is defined as "The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments". As for satisfaction, it represents the comfort and acceptability of the work system to its users, as well as other people affected by its use.

While designing an interactive product, whether it is a cell-phone, a remote control or even a calculator, one has to pay attention not only to the product's capability of carrying out their tasks, but also have in mind certain aspects, such as if it is easy enough to use, effective and enjoyable from the user's perspective. Therefore, it is critical to take into account where the product is going to be used and who is going to use it, since different populations may have different needs and requirements.

One of the major concerns while designing a product is optimizing the user's interaction with the interface provided, so that the activities that are supported can be matched by the users. This process should be iterative, parallel and incremental, involving continuous user feedback in order to develop new features or improving existing ones, as mentioned by Jakob Nielsen [5].

By supporting the development process on user's understanding, there should be some concerns regarding which users will use the product, as well as the requirements and needs the target group has. Not only do users need to perform successfully the tasks they set themselves to, the final solution must also provide an overall quality user experience to whoever uses it. In order to reach these goals, listening to what people need and getting them involved in the design allows developers to better understand their needs and goals and improve the overall usability of the product, as described by Preece et. al [2], but it should be validated through Summative evaluations and using different metrics based on user testing.

2.2 Importance of Usability

From a user's perspective, usability is crucial because it establishes the difference between successfully performing a task (with both accuracy and completeness) and failing, whether it is by committing errors and ending up frustrated, or by simply not being able to reach his goal. On the other hand, usability can affect the producer on many levels. It can mean the difference between the success and failure of the whole system, and, from a management point of view, even reduce productivity of the workforce, and costing time, effort and resources. Studies performed have shown that people prefer to use systems that provide the highest usability metrics 70% of the time, according to Nielsen [6].

If a product lacks usability (in any of the goals mentioned in the Section 2.1), the user will have trouble performing his tasks and this will cause problems of a varying degree of severity. As mentioned by Preece et al. [2], while some products, like answering machines, were considered efficient in a way it let the user carry out his tasks in a small number of steps, others, like the voice-mail system, were considered inefficient because it required the user to learn an arbitrary set of sequences for the same common task (listening to messages). Thus, a product that is not usable will ultimately cause frustration and time waste to users. On the other hand, a usable one will most likely affect positively both parties: the user and the provider. The former will not be frustrated and the enjoyment while interacting with the system will help cultivate confidence and trust in it, ultimately bringing more customers through recommendations. The latter will also benefit from reduced support costs and user errors, as well as a larger return on investment.

2.3 User Experience

As new technologies are offering increasing opportunities for supporting users with their everyday lives, this led to an expansion of usability itself, with the emergence of new events and conferences, as well as the increasing literature that is being written on the subject. Overall, the public has grown a larger awareness for this matter, and while in the past users would tolerate poor usability, nowadays it is seen as a disappointment and often a reason to not use the product again. Customers began to expect more from the product due to the increasing number of suppliers and demand usability, by being effective, efficient and satisfying as mentioned by the International Organization for Standardization 9241-11 [4] and by Ji-Ye Mao et. al [7].

With the increasing use of new technologies in various application areas by the population, such as education and entertainment, new concerns have been brought. Researchers not only have to worry about improving efficiency and effectiveness, but also have to bear in mind what the system represents and feels like to the users, in other words, the User Experience. While usability addresses how useful or productive a system is from the producer's perspective, user experience implies a more subjective nature. Hence, user experience goals tend to aim for intangibles, such as if the product is fun, entertaining or even emotionally fulfilling. Fig. 2.1 describes the different goals and the relation between them, according to [2]. The usability goals are presented in the center, which are operationalized through specific criteria. Then, in the outer circle, the more abstract user experience goals are shown.

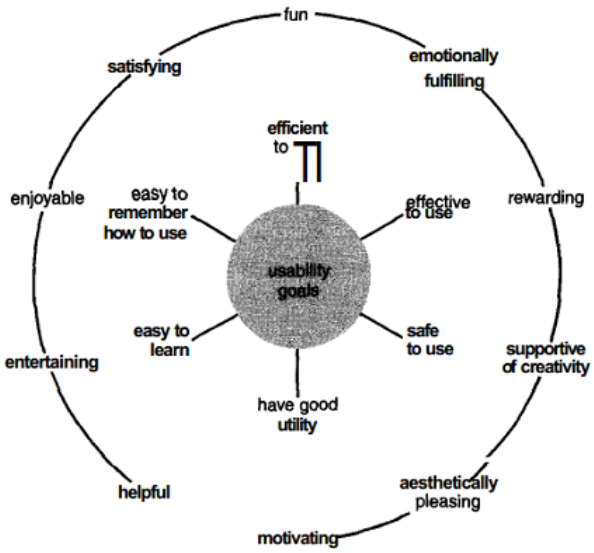

Figure 2.1: Usability and User Experience Goals [2].

2.4 User Centered Design and Iterative Approach

With the ever growing importance of the users' satisfaction while developing a product, and since there is not a clear path to reach all of the usability and user experience goals due to subjectivity, information about the users and their tasks has been taken in order to adapt the design. Therefore, it is of great importance to continuously ensure the development of the design by involving real users throughout a UCD approach, enabling a better overall understanding of the needs and goals, leading to a more appropriate and usable product.

While UCD promotes both expectation management, ensuring that the product is similar to what the user had in mind, and a sense of ownership since they participated in the design, there are some concerns. The reasons that are more consistently cited for not involving users in the development stage is the amount of time it actually takes to organize, interact, manage and control this process. It takes both time and resources to prepare interviews, meetings, workshops and user testing laboratories. There is still a debate about how actively should the users be involved in the design, as some studies have shown that largely depending on user involvement can hinder the process: Wilson et al. [8,9] presented some useful information on the topic, and concluded that users have to be educated about the design, are unaware of implementation constraints and consume precious time to the developers with meetings arrangement. Despite all these concerns, UCD is used with the sole purpose to optimize opportunities for success, providing a better final product to the users and can actually translate into money and life savings.

As mentioned by Gould and Lewis in 1985 [10], three principles were introduced for the User Centered approach to promote a useful and easy interactive system:

- Early focus on users and tasks – Meaning that the first thing a developer should worry about is identifying the users by studying their behavior and daily routines, as well as perceiving the nature of the goals they are expected to accomplish.

- Empirical measurement – By promoting interaction between users and prototypes in early development stages, performance and reactions should be observed, recorded and analyzed.

- Iterative Design – The whole process of designing a product should be iterative. When problems are found while conducting user tests, they are fixed and the new design proposal is tested once more, by observing the effects of the fixes in the overall interaction.

Hence, the methodology varies while developing a product supported by a User Centered approach. Based on *Interaction Design – Beyond Human-Computer Interaction* [2], a simple interaction design model is described in Fig. 2.2: after identifying the needs and requirements of the system, alternative designs are generated, attempting to meet the former goals. Then, interactive versions and prototypes are developed and evaluated with the target group. Based on the output given by the testers, there may be a need for redefining the needs and requirements, or just adapt the current design. If the proposed requirements are met by the prototype developed and if the results from Summative evaluation and user testing validate it, the product is ready for production stage.

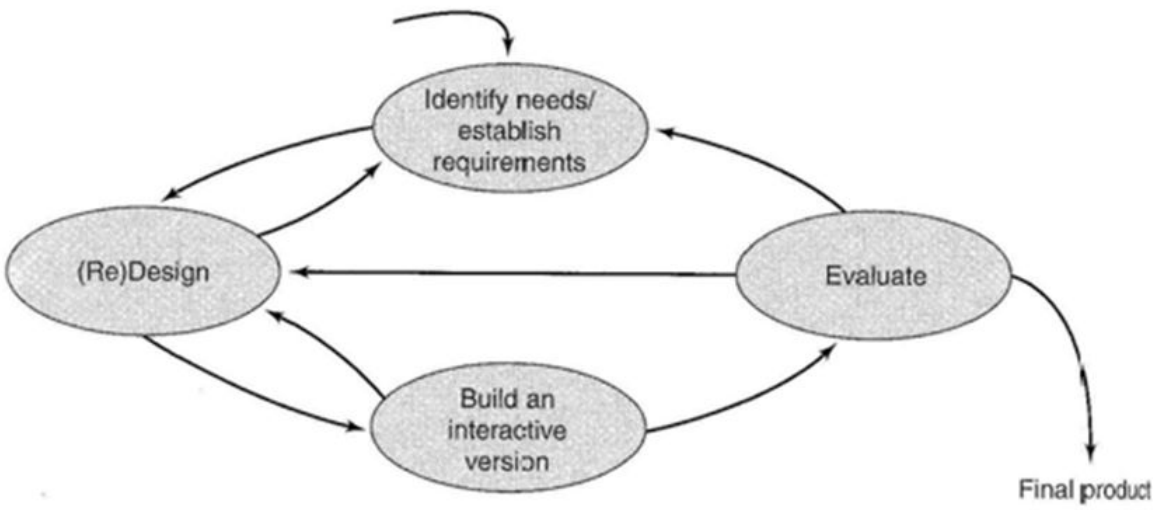

Figure 2.2: Interaction Design Model [2].

Bearing In mind that it may take several parallel designs to be evaluated throughout this cycle, or even just one at a time, and that the costs associated with user testing are not to be disregarded, the available resources available may hinder the quality of the process by reducing the number of iterations of improvement. At the time of this writing, there is still not a consensus over the number of usability

tests that should be done in order to validate a product. Nielsen [11] conducted a study on 83 of Nielsen Norman Group's recent usability consulting projects in order to analyze the correlation between the number of test users and the number of usability findings. As presented in Fig. 2.3, the small correlation shows that testing more users isn't directly associated with more insights.

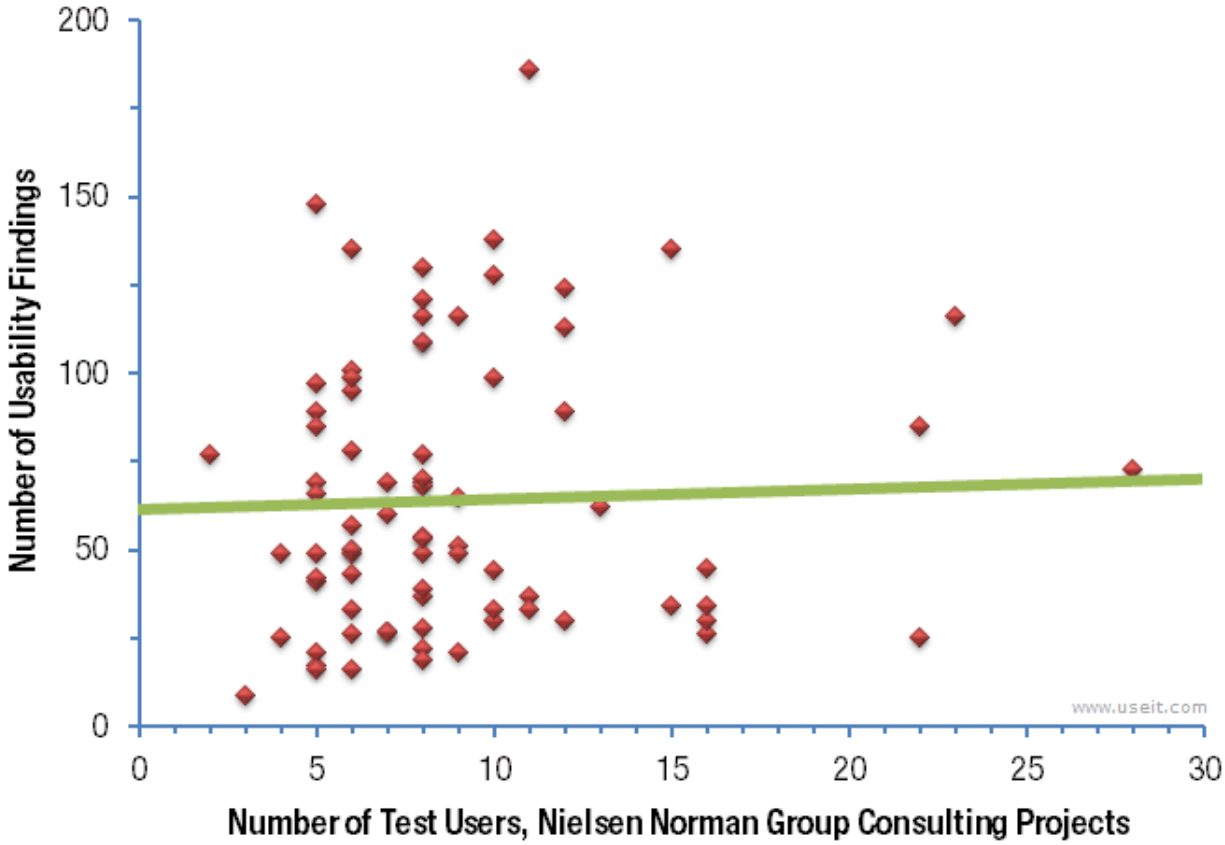

Figure 2.3: Usability findings by number of test users [11].

Nevertheless, it is implicit that the cycle will reach its pinnacle with the final product emerging in an evolutionary way, from the first prototype through to the finished project, ensuring that it provides the prescribed usability criteria.

2.5 How is Usability measured?

As concluded in Section 2.4, one of the key principles to maximize usability is by applying iterative design approaches, which progressively improve the overall quality of design from early stages. This evaluation steps allow developers to analyze the users' feedback and is continuously used until the system reaches an acceptable level of usability.

The favored method to ensure the quality of the design is testing with actual users on a working

system, requiring focus on the end-users' needs and adapting the design based on them. Some factors can hinder this approach, such as the users' schedule and the budget of the development team, reduce its proficiency. There are other alternatives, such as user testing on prototypes (both low and high fidelity ones), cognitive modeling or a usability audit conducted by experts using heuristic evaluation. On the following sections, both user testing and heuristic evaluations are described and further explained their processes, as these techniques will be used during the development of the proposed solution.

2.6 Heuristic Evaluation

Usability evaluation has been around since the 1980s, where similar techniques have been used for this purpose. In heuristics evaluation, experts are guided by heuristics and simulate what an actual user would do, by stepping through tasks and thus identifying the major problems with the system. Compared to user testing, this approach is far less expensive and does not have the problem of gathering a group of users nor preparing a laboratory. After experts detect a list of problems and conditions, feedback is reported to the developers in order to improve the interaction with users.

According to Jakob Nielsen's article back in 1995, heuristic evaluation had its first steps when the heuristics were originally developed in collaboration with Rolf Molich back in 1990 [12]. They were given this name, heuristics, because they represent broad rules of thumb and not specific usability guidelines, although they resemble some high level principles of design. Since then, Nielsen refined the heuristics based on a factor analysis of 249 usability problems (Nielsen, 1994a [13]), focusing on a set of heuristics that provided maximum explanatory power, resulting in the set of 10 revised heuristics (Nielsen, 1994b [14]), only this time including the questions addressed while assessing each one of them:

- Visibility of system status: Are the users kept informed about what is going on? – The system should provide appropriate feedback within real time and inform users of the current state.
- Match between system and the real world: Is the language used at the interface simple? – The system should use concepts familiar to the user rather than specific system-oriented ones, and make information appear in a natural order.
- User control and freedom: Are there ways of allowing users to easily escape from places they unexpectedly find themselves in? – Users often perform errors and need an easy option to undo the mistake, avoiding a large set of steps.
- Consistency and standards: Are the ways of performing similar actions consistent? – The system should follow known conventions and avoid situations where users wonder if different words and situations mean the same thing.

- Help users recognize, diagnose, and recover from errors: Are error messages helpful? – Error messages should be expressed in plain language and easy to understand, as well as precisely indicate the problem and the best solution available.
- Error prevention: Is it easy to make errors? If so, where and why? – The system should always try to prevent errors from occurring, either by eliminating error-prone conditions or by using confirmations options before forcing the user to commit to actions.
- Recognition rather than recall: Are objects, actions and options always visible? – The user should not have to remember information about previous actions, thus, the instructions must be visible and easily retrievable whenever appropriate.
- Flexibility and efficiency of use: Have accelerators been provided that allow more experienced users to carry out tasks more quickly? – Usually missed by the unexperienced users, shortcuts provide ways for expert users to optimize their actions and to tailor the most frequent ones.
- Aesthetic and minimalist design: Is any unnecessary and irrelevant information provided? – Information that is irrelevant for the current state of the system competes with the actual useful information the user needs, therefore reducing its visibility.
- Help and documentation: Is help information provided that can be easily searched and easily followed? – Even though it is preferable that the system can be used without any documentation, users may need help, and such information should list the concrete steps to solve a specific task.

In the context of the presented thesis, evaluation represents a major factor for any Information Visualization work, and Heuristic Evaluation is one of the most used and reliable method for this purpose. Various sets of heuristics have been proposed, including the set of 10 revised heuristics by Nielsen in 1994 [14], but the lack of consensus on which one to use represented a problem for the **InfoVis!** (**InfoVis!**) community. A study was performed by Forsell [15] in 2010 to empirically determine a new suitable set of heuristics in order to better evaluate interactive visualizations, resulting in a synthesized set of 10 heuristics with the highest explanatory coverage for 74 different usability problems.

During the study presented in Chapter 4, and due to the nature of the work, Forsell's [15] synthesized set of heuristics was applied, bearing in mind that it is suited for Information Visualization purposes and that it provides a wide explanatory cover of usability problems associated with InfoVis.

Following is presented a set of heuristics developed in 2010 by Forsell [15] that suits Heuristic Evaluation on InfoVis and that will be used throughout this document in order to continuously evaluate the solution's development:

- B5. Information coding. Perception of information is directly dependent on the mapping of data elements to visual objects. This should be enhanced by using realistic characteristics/techniques

or the use of additional symbols.

- E7. Minimal actions. Concerns workload with respect to the number of actions necessary to accomplish a goal or a task.
- E11: Flexibility. Flexibility is reflected in the number of possible ways of achieving a given goal. It refers to the means available to customization in order to take into account working strategies, habits and task requirements.
- B7: Orientation and help. Functions like support to control levels of details, redo/undo of actions and representing additional information.
- B3: Spatial organization. Concerns users' orientation in the information space, the distribution of elements in the layout, precision and legibility, efficiency in space usage and distortion of visual elements.
- E16: Consistency. Refers to the way design choices are maintained in similar contexts, and are different when applied to different contexts.
- C6: Recognition rather than recall. The user should not have to memorize a lot of information to carry out tasks.
- E1: Prompting. Refers to all means that help to know all alternatives when several actions are possible depending on the contexts
- D10: Remove the extraneous. Concerns whether any extra information can be a distraction and take the eye away from seeing the data or making comparisons.
- B9: Data set reduction. Concerns provided features for reducing a data set, their efficiency and ease of use.

During the Heuristic Evaluation process, the evaluator, preferably an usability expert, proceeds to analyze the given prototype and reporting any problems detected on a list with a reference to the violated usability principle. Furthermore, the evaluators are asked to rate the severity of the detected problems, using Nielsen's four step scale [16], from 0 = I don't agree that this is a usability problem at all, to 4 = Usability catastrophe: imperative to fix this before product can be released.

2.7 User Testing and Usability Metrics

Measuring the quality of a design by comparing users' performance is the preferred method, assuring that, although using a relatively small sample size, the difficulties those testers had can potentially

translate to the actual problems the clients may have in the future. These tests are usually conducted in controlled conditions and involve the selected users to perform well predefined tasks, involving and testing every mechanism associated with the product. After that, the data collected is analyzed and used to resolve and therefore improve problems in the system.

User testing is used so that the product can be tested and experimented by a group of users from the focus group, and to guaranty that it is usable by the time it is released to the clients. As mentioned in the previous paragraph, the conditions of the tests are controlled, meaning that the tasks the user will perform were clearly defined before, and that the expected amount of clicks/seconds and number of errors were also estimated. Thus, every piece of information that can be used from these tests is collected: the time taken to complete a task, the number of errors, the type of errors and sometimes the path chosen by the user. To make sense of this information, observational data, recordings, user satisfaction questionnaires and interviews are also used for context.

When performing user testing, developers face the problem of defining which metrics to be evaluated and how each one of them is going to be measured. As mentioned by the International Organization for Standardization 9126-4 [17], three basic usability metrics must be included, and those are effectiveness, efficiency and satisfaction. There are many ways to measure these as it depends on the discretion of the evaluator and on the conditions of the laboratories used. Following are presented the most common ones.

Effectiveness represents the accuracy and completeness with which the users achieve the given goals, and can be further evaluated by measuring the completion rate and the number of errors each one of them had. When a user successfully completes a task, a binary number '1' is assigned to it and '0' if not. When evaluating the number of errors, a short description of the situation (whether it is a slip, an unintended action or a mistake) should be assigned, giving developers a powerful diagnostic tool to avoid the possibility of replicating such events.

Efficiency is related to the resources expended by the users while performing given tasks, and it can be measured in terms of the time taken (seconds) or even by the number of clicks or taps used to complete the goal. Preece et. al acknowledges [2] that the more resources are used while doing a task, the higher the chance the user has to drop it due to frustration.

Satisfaction has a subjective nature and therefore is measured through standardized satisfaction questionnaires, which can be administered after each task or in the end of the testing session. Test level questionnaires are typically used right after a task is completed and measure how difficult it was, complementing important task-performance data, such as time and completion rates, Sauro suggests [18]. In the end, a test level questionnaire is used to measure the user's perception of the overall ease of the system, being the SUS the most popular standardized one, accounting approximately for 43% of post-test questionnaire usage in a collection of unpublished usability studies, as mentioned by Sauro

and Lewis [19].

In 1996, John Brooke published an article mentioning the concept of a “System Usability Scale”: a “reliable, low-cost usability scale that can be used for global assessments of systems usability.” [20]. The SUS provides a ”quick and dirty”, while reliable, tool for measuring satisfaction on users. This type of questionnaire is delivered generally after the test is taken and consists of a set of 10 different questions with five response options for each respondent; from ”Strongly agree” to ”Strongly disagree”. After the calculation, the result provides an overall usability and user satisfaction index, ranging from 0 to 100 that should not be associated to a percentage. SUS was introduced with the same strengths that as made it an industry standard; due to its inexpensiveness, how easy it can be administered to participants and its usefulness on being used in small sample sizes, while providing reliable results. For these reasons, the authors decided to use this mechanism to evaluate users’ satisfaction on the developed Information Visualization (infoVis), which will be explained in detail in Chapter 5.

2.8 Other Applications of Usability

There are many areas where usability is used. As mentioned throughout this Chapter, Human Computer Interaction (HCI) represents a major area where the concept of usability is applied, since it involves perfecting the connection between people and machines, the major contribution of usability is in Human-Computer Interaction, by improving systems and making them easy and enjoyable to use by the final users. Human factors psychology, also known as ergonomics, is a vast field which discovers and applies information about human behavior, abilities and limitations to the design of products, systems, jobs and tools, making them safe, productive and comfortable for use. Psychology is applied to study human perceptive and cognitive processes to improve and adapt technology or equipment to better suit human capabilities. Some studies developed in this area include evaluating the usability of gadgets [21] or redesigning health care systems [22].

Research has been made in many other areas as presented by Maria Riveiro et al. [23], usability evaluation was used to assess visualizations of normal models built from representative data. In this article, sea surveillance systems are addressed, as well as their approaches using statistics to detect irregular patterns and consequently suspicious behavior. Since these systems typically generate high false alarm rates due to both the size of the data and the complexity of such models, visualizations of this information were developed and usability testing was used to better comprehend and validate them. User testing was based on a list of representative tasks, such as classifying vessels or identifying passenger routes.

3

Related Work

Contents

This study on visualizing usability emerges from the intersection of research on infoVis and UCD. Metrics can be recorded by performing user testing and an overall assessment of the system usability can be derived from heuristic evaluations, but the authors believe that information visualization can provide an interactive way of expressing and analyzing such data, as well as representing a strong predictive tool that can improve the overall design. While there is plenty of work developed in the area of information visualization, there is a lack of research relating it to usability metrics, which hindered the investigation on related work.

Although there are some tools already in the market that tend to analyze user data, such as Google Analytics [1] or Hotjar [2], but their focus is centered on the analysis of large scale usage data through techniques such as plots, bar and pie charts, and in the latter's case, heatmaps. Thus, these tools diverge from what this study envisions, providing analytics on large scale data more focused on the visitors' usage of a given website. On the other hand, the proposed study intends to analyze user data on a more preliminary phase; and to specifically rely on data gathered from user testing sessions, in order to further provide information about derived metrics based on effectiveness, efficiency and satisfaction, as well as identifying troublesome areas on the product before it reaches the production phase.

Since this thesis intends to study and analyze an aggregation of both objective (efficiency and effectiveness metrics) and subjective data (satisfaction metrics), an analysis was done in this chapter regarding both types mentioned. During the analysis of related work, the visualizations found were categorized into three different types: pure Statistical Analysis, Correlations and Performance/Benchmarking. The main goal of this analysis was to realize how each type of visualization techniques used diverge from the other ones and to further assess each work, identifying potential flaws and strengths of each. The analysis performed in this chapter should provide both guidelines and experience regarding mechanisms and techniques that should be used and ultimately, to improve the overall quality of the presented work.

3.1 Statistical Analysis

The importance of Information Visualization in statistical analysis has been increasing steadily following Tukey's Exploratory Data Analysis [24]. A comparative study between InfoVis and Statistical Analysis developed by Gelman et. al [25] suggests that there is lack of cooperation between both fields, and that by joining InfoVis' capacity of attracting people's attention with the conveyance of data provided by pure statistics, both areas could improve their visualizations.

On the field of educational data, plenty of studies were performed in order to optimize the learning process for both students and teachers. There are some visualizations on students' participation and

[1] Google Analytics Solutions. Available at https://www.google.com/analytics/ [Consulted at 05/03/2018]
[2] HotJar. Available at https://www.hotjar.com/ [Consulted at 05/03/2018]

performance levels, as well as on structured education content that were developed since the beginning of the current decade. Tervakari et. al [26] and Hassan et. al [27] focused the presentation of information mainly targeting students' performance and attendance levels, while Chen [**?**] relied more on a predictive study and a possible dropout identification between groups of students. After analyzing the visualizations and the feedback given by user testing, the main objective, which was to promote self reflection and awareness, was sometimes misunderstood and the usefulness was questioned by the students, due to the density of the presented data (Fig. 3.1 and Fig. 3.2), and that specifically in *DropoutSeer* by Chen [28], the identification of risk factors was affected by the quality of the predictive model used.

The following figures present visualizations developed by Tervakari et. al [26], in the context of statistical analysis of students' performance and attendance levels. Fig. 3.1 describes an interactive bundle visualization, representing relationships between students, study topics, web resources and discussions they participated in. Fig. 3.2 shows an interactive visualization dashboard displaying activity and participation levels of student groups.

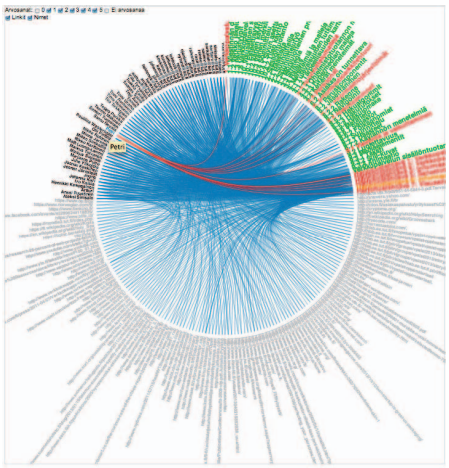

Figure 3.1: Students, study topics, web resources and discussions relationship Visualization [26].

Figure 3.2: Activity and participation levels of student groups [26].

The brazilian population's voting behavior was also analyzed in *CivisAnalysis* [29], where a set of visualizations was implemented in order to represent different political positions and dimensions. Expected InfoVis behavior was used, from filtering by parties, states or deputies, and other data exploration features. The overview controlled the majority of the interactive behavior of the visualization through a timeline based both on period and party selection. The nature of the data used hindered creativity: some of the views implemented consisted of Scatter plot visualizations with color differentiation, causing some problems in the exploration phase, particularly in focusing on smaller parties. In the final analysis of the work developed, the author Francisco Borja acknowledges that correlation tools could be further developed, and mentions as example the possibility of relating roll calls of the same legislative motion over time. The following Fig. 3.3 shows *CivisAnalysis'* [29] Inspection module presenting the Brazilian Chamber of Deputies in the year 2005-2006, after being selected in the timeline of legislatures at the bottom. In the same figure, there can be observed different areas: (*A*) shows the distribution of deputies per party, while (*B*) presents the political spectrum of deputies and (*C*) shows the political spectrum of roll calls.

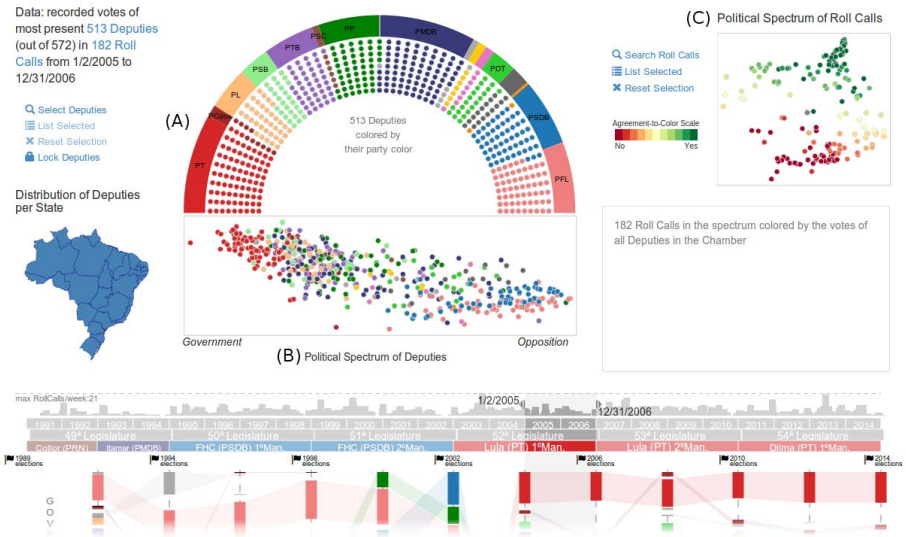

Figure 3.3: Inspection module visualizing the Brazilian Chamber of Deputies in the year 2005-2006 [29].

3.2 Correlation

A correlation is defined as a statistical link that can be causal or not, between two different variables or attributes. Some visualizations tend to explore this characteristic of data, in order to provide evidence and identify relationships that can be exploited in practice.

Answering questions about complex relationships between multi tabular data requires a statistical analysis technique to make them visually discernible. Kamasan et. al [30] introduced a study to visualize statistical tests and to display the distributions of data using color schemes, providing the user a tool to quickly evaluate large sets of data for statistical correlations. As shown in Fig. 3.4, a visualization dashboard developed by Kamasan et. al [30] shows four different views based on correlations between attributes and their statistical formulas: the desired attributes are selected in (*A*) and the relationships between them are highlighted by both a Heat-map (*B*) and a Relationship visualization (*C*). Since the main goal of the presented work is to improve the analysis of performance related metrics, these visualizations provided considerable value to the current study, although the lack of interactivity represented amongst them is unwanted.

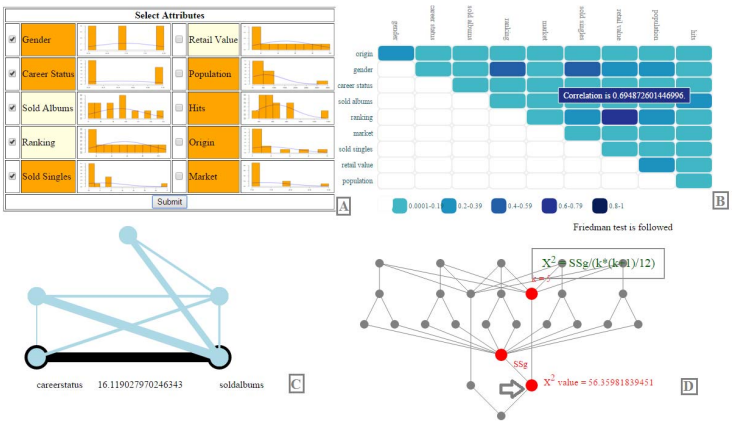

Figure 3.4: Visualization dashboard with correlations between attributes and their statistical formulas [30].

Sandra Cano et. al [31] developed an interesting study on visualizing structures, patterns and correlations in the context of Spanish pronunciation. The quality of the input signal is compared with the training done on the vowels and these differences and similarities are highlighted through the use of Chernoff faces [32] and diagonal lines. As shown in Fig. 3.5, a views model presents information regarding incorrect spanish pronunciation of the *E* vowel. The first view (Fig. 3.5, View 1) represents characteristics obtained through the use of Chernoff faces [32]. The second view (Fig. 3.5, View 2) measures similarities in pronunciation characteristics between the test signal and the training one. The third view (Fig. 3.5, View 3) uses a Self Organizing Maps technique and positions groups of vowels based on their characteristics similarities. The final view (Fig. 3.5, View 4) is represented by a diagonal line indicating the correlations between the five vowels in Spanish and the test signal. According to the evaluation study performed in the same work, it was concluded that some views presented too much information for the user to pay attention at a time (Fig. 3.5, view 2), while others were did not present its information in a clear way (Fig. 3.5, views 1 and 4), causing doubt amongst users. Although the user was able to save representations and choose the ones to view at a time, no sort of interaction between the visualizations was provided. As for other areas regarding correlation studies, Wei Zeng et. al [33,34]

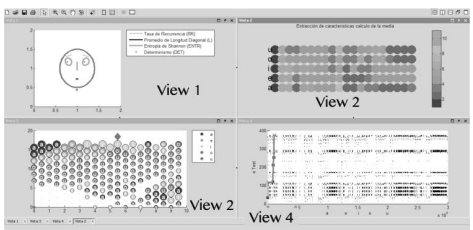

Figure 3.5: Views model regarding incorrect pronunciation of the E vowel [31].

did extensive research concerning human mobility and interchange Patterns, with the objective of better understanding the motivation behind people's movement.

Zeng produced three contrasting studies in 2017 [33] with the main goals of providing an overview

of the areas of interest in Singapore and to analyze the connection between people's movement and the points of interest signatures in different categories. Fig. 3.6 shows a visualization developed by Wei Zeng et. al [33] regarding exploration and comparison of Point of interest mobility signatures in different categories: *a)* university, *b)* factory, *c)* shopping center, *d)* residence and *e)* airport. The reliance on signatures provided an easier and visually appealing way to recognize movement peaks and mixed usages, but scales poorly, as each one occupies certain space and thus the number of representations is limited. Furthermore, expert evaluation revealed that more user interaction should be implemented and that the information retrieved may be biased because the usage is restricted to public transportation data.

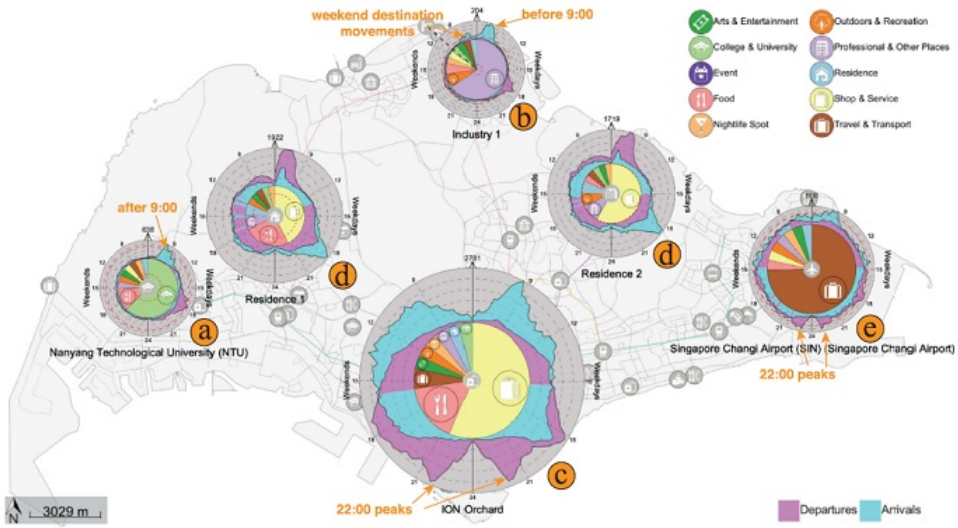

Figure 3.6: Visualization with exploration and comparison of Point of interest mobility signatures [33].

Zeng had also presented work where the formulation considers the study of interchange patterns at different scales back in 2013 [34]: from train stations in a metro system, to crossroads and regional zones in a city, while using visualization approaches such pattern extraction and data aggregation. This paper presents a novel method of visualizing and exploring interchange patterns named *Circos Diagram*, based on an interchange matrix that summarizes the flow volumes of different possible routes across a junction node. With the help of interactive operations, such as time controlling and "roll out" actions, the user can better examine temporal variations and draw conclusions concerning traffic peak hours.

Human control related studies include traffic management, and Chris Musialek [35] presented *Traffic Trender*, a systematic tool to identify bottlenecks, which combines a zoomable Tree-map with line charts and filtering functions, allowing the user to find trends in traffic incidents, both geographically and over-time. Besides identifying patterns, its use is intended to provide the ability to easily detect outlier and anomaly information through the cross reference available between both views.

3.3 Performance and Benchmarking

The assessment of performance related metrics has been commonly obtained through the analysis of simple visualization tools as graphs and bar charts, such as Price's [36] and Gusev's [37] work. In 2016, Price et. al [36] studied and examined whether the use of an info-vis could lower working memory demand on elder people and if it would consequently affect complex decision making in a positive way, while displaying the results through bar charts. On the other hand, Gusev et. al [37] presented in 2014 an extension of their *EDUCache Simulator*, which consisted in graph based visualization of the cache performance behavior when a program accesses huge data arrays by analyzing real three level cache systems.

The authors believe that Information Visualization represents a powerful tool to help displaying such performance measures, as well as easing the process of understanding, and therefore lowering the cognitive workload associated with the information that is presented.

Mamani et. al [38] developed in 2014 a visualization tool to help providing measuring software performance. In this paper, and alternative to the commonly used tree widget is presented: a sunburst-like visualization to represent program executions and time distribution through color maps. Although this sunburst visualization provides more interaction than other performance related works, it still lacks innovative usage of these mechanisms, providing basic tool tips, source code browsing through clickage and highlighting component-binded methods. Mamani mentioned that the work is still in an early phase, and that its hypothesis is to enable faster extraction of relevant information, as well as identifying possible execution bottlenecks. M. Weber et. al [39] has also done work on the detection of performance variations and application bottlenecks and presented an approach to facilitate visual analysis of performance data. As expected, performance related visualizations, as the ones described on the previous paragraphs, are associated with losses both on user interactivity and quality of the techniques applied, and while the nature of the data used is applied to a specific target group of users, infoVis represents a powerful tool to display this kind of information.

As presented in the following Fig. 3.7, processes and their respective runtimes are displayed through a Runtime variation visualization analysis, developed by M. Weber [39]. Several processes are shown in the middle with higher runtimes in their dominant function, which can be perceived by color usage.

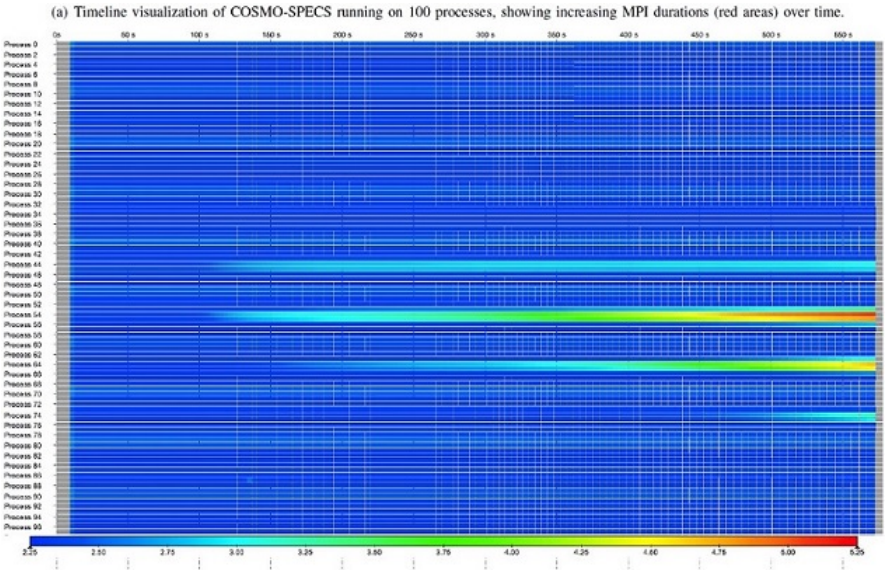

Figure 3.7: Runtime variation visualization: processes and respective runtimes [39].

3.4 Discussion

The related work presents visualizations which can help contributing with additional information about the way the solution will evolve. Therefore, after studying each work previously mentioned and the visualizations that they implement, a classification was made for each work, in order to better classify and analyze them. Based on whether those visualizations can or can not provide certain functions and characteristics that the authors believe are essential while presenting data through Information Visualization, the following characteristics were used for the purpose of this discussion:

- Clearness - Represents the objectiveness of the visualization and the ability to provide the intended information to the users and if they can take conclusions from it. The assessment of this characteristic is preferably done through user's feedback, due to its subjective nature.
- Exploration - The ability to provide simple exploratory functions, such as filtering, zooming and tool tip information.
- Correlation - The ability to provide enough information to correlate attributes and to find patterns.
- Interactivity - Represents to ability to allow users to interact with the visualizations and have some control over it.
- Scalability - Represents the ability to represent larger sets of data without compromising the quality of the visualization presented.

Table 3.1: Comparative study on related work visualization types.

Name	Type	Clearness	Exploration	Correlation	Interactivity	Scalability
Tervakari [26]	STAT	NO	YES	YES	YES	NO
Chen [28]	STAT	NO	YES	YES	YES	YES
Hassan [27]	STAT	YES	YES	YES	NO	NO
Borja [29]	STAT	YES	YES	NO	YES	YES
Cano [31]	CORR	NO	NO	YES	NO	NO
Vaishnavi [30]	CORR	YES	YES	YES	NO	YES
Zeng [33]	CORR	YES	YES	YES	NO	NO
Zeng [34]	CORR	YES	YES	YES	YES	NO
Musialek [35]	CORR	YES	YES	YES	NO	YES
Price [36]	PERF	YES	NO	NO	NO	YES
Gusev [37]	PERF	YES	NO	NO	NO	YES
Mamani [38]	PERF	YES	YES	NO	YES	YES
Weber [39]	PERF	NO	YES	NO	NO	YES

The presented Table 3.1 provides a comparison of the visualizations studied and mentioned in the related work based on the following characteristics: Type(of study), Clearness, Exploration, Correlation, Interactivity and Scalability. Based on this information, the authors were able to extract information and conclusions from the three types of visualizations studied.

First of all, statistical visualizations provide great tools that allow exploring, correlating information and interact with appealing techniques. Some of the views used over the assessed work include bundle, network, clustering and tag clouds visualizations, but they commonly lack objectiveness when dealing with large datasets and thus have problems with scaling. On the other hand, performance visualizations present contrary characteristics, where its simple techniques, such as graphs, bar charts and time line based visualizations compromise user control functions, while accessing clearness and scalability. Finally, the correlation visualizations showed mixed results, affording lower interactive power compared specially to statistical ones, and presenting scaling vulnerabilities while affording great tools to compare and connect information through necklace, density and heatmaps.

Regarding the comprehensive study of the related work alongside its visualizations and respective classifications, *CivisAnalysis* [29] succeeded and brought together all of the characteristics that were analyzed in the discussion but correlation, doing this while providing positive aesthetic levels through its techniques' appeal. Francisco Borja mentioned that, understandably, the visualization could be further optimized, from adding more dimensionality reduction techniques, to improving exploration and relation phases within it.

The type of data used in *CivisAnalysis* and in the other studies of the related work is obviously different from what this thesis intends to handle. Besides that, the authors stress that the nature of this thesis' solution proposal is innovative and that there is not much work nor studies relating User Experience and Information Visualization, thus the lack of absolute correspondence between the related work and the presented solution. Overall, the authors acknowledge *CivisAnalysis* and believe it can represent an important cornerstone throughout the development of the final solution, bearing in mind

the former's weaknesses and strengths.

In order to provide a novel solution based on a dashboard with different Visualization techniques and to better explore Usability, the analysis of the related work along with each type of visualization contributed to the development of the presented thesis in a sense that it allowed the authors to better understand both the vulnerabilities and advantages of each visualization technique used. The related work allowed an improvement on the decisions regarding the types of visualizations implemented in the solution. The main challenge resided in channeling these conclusions in a way that could help provide not only an innovative experience but also contributing to usability metrics analysis related tasks, and thus develop a reliable proposal.

The following Chapter explores all the mechanisms used throughout the development phase, along with the evaluation steps that guided its progress. The importance of the related work research is reflected on the choices made on the prototype design, along with the relevance and goals the authors felt each Visualization was intended to display.

4

Proposed Solution

Contents

In this chapter, a solution proposal to the presented issue is presented, which is visualizing and exploring new ways of analyzing and presenting Usability related metrics, along with its adjacent user tasks. All mechanisms used and phases that supported the development of the final visualization are described by the authors, as well as some reserved space to address several problems and decisions and how those were overcome.

From the development approaches adopted to the final phase of the visualization, other important aspects are properly detailed throughout the following sections. User requirements obtained through surveys and interviews; the initial set of questions for the visualization; decisions made regarding the dataset used; the rationale behind the Architecture and Plan; the different phases of the presented prototype, as well as the formative evaluation steps used throughout the development are discussed.

4.1 Development Approaches

As mentioned in Chapter 1 and further highlighted in Chapter 2, the progress and evolution of the visualization, which envisions to facilitate the extraction of information regarding usability metrics, is based on the feedback given by the potential users of the product - which is typically used to identify usability problems with the prototype in early stages, so they can be fixed before the final iteration of the design. Based on this description, a focus group of six elements was used, allowing continuous judgment on the quality of the prototype and providing information and potential resolutions for future improvements, through discussion sessions and formative evaluations. The focus group was restricted to people with minimal experience both with Information Visualization designing and with Usability related approaches, whether by attending corresponding courses, or by daily work experience.

The design provided in this solution is based upon an explicit understanding of users, tasks and environments, and involves them throughout the study, by following three UCD principles already mentioned by Gould and Lewis in 1985 [10], considered stepping stones to improve the quality of the final iteration of our prototype:

- Early focus on users and tasks – There should be a phase of requirements gathering and identifying the users' needs as well as understanding the nature of the goals he is trying to fulfill.

- Empirical measurement – By promoting interaction between users and prototypes in early development stages, performance and reactions should be observed, recorded and analyzed.

- Iterative Design – The whole process of designing a product should be iterative. When problems are found while conducting user tests, they are fixed and the new design proposal is tested once more, by observing the effects of the fixes in the overall interaction.

Thus, before reaching advanced states of the design, the authors considered vital to first reach out to the target group and understand its needs and requirements. It is important to model the design according to the expectations of the users, while managing the ability for them to conclude their own daily tasks [2]. In order to capture the characteristics of the intended user group, and to better conceptualize a initial model, series of interviews and Surveys were used in the early stage of the solution progress. From the first non functional prototype of the conceptual model until the final functional prototype, every phase went through a formative evaluation process to check if it met users' needs: informal discussion sessions and heuristic evaluations were the most common approaches, which guided the state of the design to continuously improve through new design proposals.

The following sections further describe the evolution of the presented solution and how each step influenced our decisions regarding the design.

4.2 Requirements Gathering

Although the authors had an early vision of the design concept they proposed to present, that idea of what the solution would develop into could diverge from what the target group would expect. In this line of thought, understanding what the product under development should do and ensuring that it supports the target group's needs was considered a priority, being indispensable in any UCD approach and, more specifically, to Design Thinking.

Usability experts and test conductors, as well as people who want to analyze usability related metrics represent the target group for this study, and in order to better understand their current tasks and the ones they intend to perform with the help of infoVis, the authors conducted a series of Interviews and Surveys. The user requirements gathering aimed for students and professors related to both Usability and Information Visualization areas, as well as people working in User Experience.

Furthermore, a initial set of Questions is derived from user requirements, which will guide the development of the presented solution, based on the tasks and pieces of information users would like to obtain, opposed to what they currently do with user testing data.

4.2.1 Surveys and Interviews

Surveys and Interviews were used in order to retrieve information about the way users expect to interact with a given design, and therefore it was important to organize a study before developing the first prototype. There were some concerns regarding the metrics that the users wanted to analyze, their priorities of exploration and mechanisms they would like to complement the analysis of these metrics. Besides, the authors wanted to confirm and verify how useful an infoVis would be in the context of studying usability related data.

A survey was handed to 67 people, where 56.7% of the population's age ranged between 18 and 24 years old, with the other 43.3% with ages comprehended between 25 and 44 years old. The grand majority were male (80,6%) and all had at least a bachelor degree in computer science. Furthermore, these studies were restricted to people who had at least completed a Human Computer Interaction course, and therefore any prior knowledge about Usability. There was an effort to obtain feedback from users about how important Usability is in their opinion, and how it could be complemented with the help of Information Visualization.

From the inquired, 73.1% said that Usability is of great importance during the development of a product, classifying its importance from 0 to 5 with 5. The most sought after objectives of Usability according to the users are effectiveness (75%) with efficiency and satisfaction coming at a close second (62% each). After the three already mentioned, utility was also considered an important objective contributing to an usable product, with 52% of the responses mentioning a great importance. The common consensus was that SUS post questionnaire was a well known tool for satisfaction measurement, as 98% of the inquired said they were familiarized with it, while 72% of them considered of great importance the measurement of satisfaction during product testing. The authors proposed then several Usability metrics for effectiveness, efficiency and satisfaction measurement, and it was possible to conclude that the most important ones for users were number of errors and satisfaction level (69% classified it as of great importance); completion time (61%); number of frustrations (50%), percentage of tasks completed (45%) and time spent on errors (39%).

When asked, Information Visualization was mostly regarded as a very important tool by the majority of the responses (63%) for Usability data analysis. Furthermore, the authors tried to gather additional insight about the characteristics users prioritize for an infoVis: the results concluded that on the Usability context, Clearness, Consistency, Interactivity and Exploration were more important for users than characteristics like aesthetics and scalability. The users were also more interested in mechanisms such as comparative analysis (94%) and correlation of metrics (77%), as well as identification of clusters and groups (56%). There was also some interest regarding the visualization of statistical derived metrics, such as time spent per section, mean values, standard deviations, errors over time and percentage of the population with a task completed over time. Besides, some suggestions included the identification of troublesome areas and exact location of errors.

After the initial Survey results, the authors interviewed three Usability experts that worked on User Experience (UX) and asked them how they perceive the application of infoVis on the particular context of Usability, and what time of information and conclusions they expected to retrieve through the analysis of such tool. As concluded through Surveys handed, the interviewed were mostly interested in the the most common measures for effectiveness, efficiency and satisfaction, and referred the significance of correlating such measures. Naturally, the importance of prioritizing information and automatizing pro-

cesses of exploration was mentioned, such as tracking ”the most affected area” either by cursor usage or errors made, identifying outlying behaviors or comparing different groups of test results, since the use of infoVis ”could ease the analysis of such metrics”. One interviewed, who lectures UX related courses, mentioned the results from user testing can sometimes be exhaustive to understand and such tools could reduce cognitive workload associated with it, and how important it was ”reaching the ’truthfulness’ behind the data”. Furthermore, other features such as a comparison tool for different versions throughout the development of a product could present additional insight regarding the improvements made on each stage.

4.2.2 Initial set of Questions

The results gathered from the processes mentioned in Section 4.2.1 had substantial value for the presented study, as it provided insight regarding what information the infoVis was going to show and consequently allowed the design of the first low level prototypes.

Bearing in mind the feedback given by the target users, the authors completed the first set of ten Questions that should be answered through the visualization developed:

- Are time consuming areas associated with more errors?
- Is the area that takes users the most time also the most error prone?
- Do users make more errors when trying to spend less time on a given section?
- What is the average time taken in each section?
- What areas in the interface take the most time/cause the more errors?
- Does SUS rating correlate with number of errors?
- Does SUS rating correlate with completion time?
- Do tests with similar metrics(time and number of errors) share similar SUS ratings?
- Which metrics affect SUS the most?
- How much does it take until *2/3* of the users complete a certain task?

These questions would present themselves useful in the creation of the first conceptual idea of the design, since every visualization and interaction depends on what questions are going to be answered through the design. The set was later revised during the low level prototype design, as discussion sessions led the authors discard particular mechanisms introduced. This segment will be further described on Section 4.6.2.

4.3 Proof Of Concept Methodology

Since the main goal is to assess usability through the potential of interactive visualizations, a large enough dataset would have to be used in order to draw conclusions and extract correlation patterns. We defined that one of the requirements was that the data would have to provide enough information about efficiency, effectiveness and satisfaction, inferring not only from metrics such as time, number of errors and SUS, but also from derived ones.

For the solution, and more specifically to gather necessary user data, it was initially proposed the development of a web browser plugin which would allow the tracking of the user's interaction with a given interface and test with a considerable amount of users to retrieve enough data to begin prototyping.

The authors felt that such solution would diverge from the main goal of this thesis and consume unnecessary time just for data retrieval. As the main objective is to study Usability measures assessment through infoVis, and in order to allocate more time for acUCD and the Iterative Design of the solution, the authors decided to apply a Proof of Concept (POC) methodology. Since the veracity of the information displayed through the developed infoVis was not important to demonstrate the feasibility nor the practical usage of the solution, and since the conclusions that the authors were aiming for did not have any relation with the dataset used, POC was seen as a viable methodology to surpass the given problem.

The main goal POC usage is to quickly and efficiently find solutions to technical problems, such as how systems can be integrated or, in this particular case, how a given interface display would interact with users, without compromising the presented study.

4.4 Architecture and Reasoning

Before beginning with prototyping, it should be taken in consideration how the solution will be developed, as well as identifying which components build up for it. This section is reserved for the analysis of how the solution architecture is built and how each mentioned component interacts with each other.

As mentioned throughout the document, the presented solution revolves around the concept of UCD, and therefore the architecture shown in Fig. 4.1 reflects the methodologies used to accomplish that goal, representing a concept model regarding the whole logic and architecture of the proposed solution.

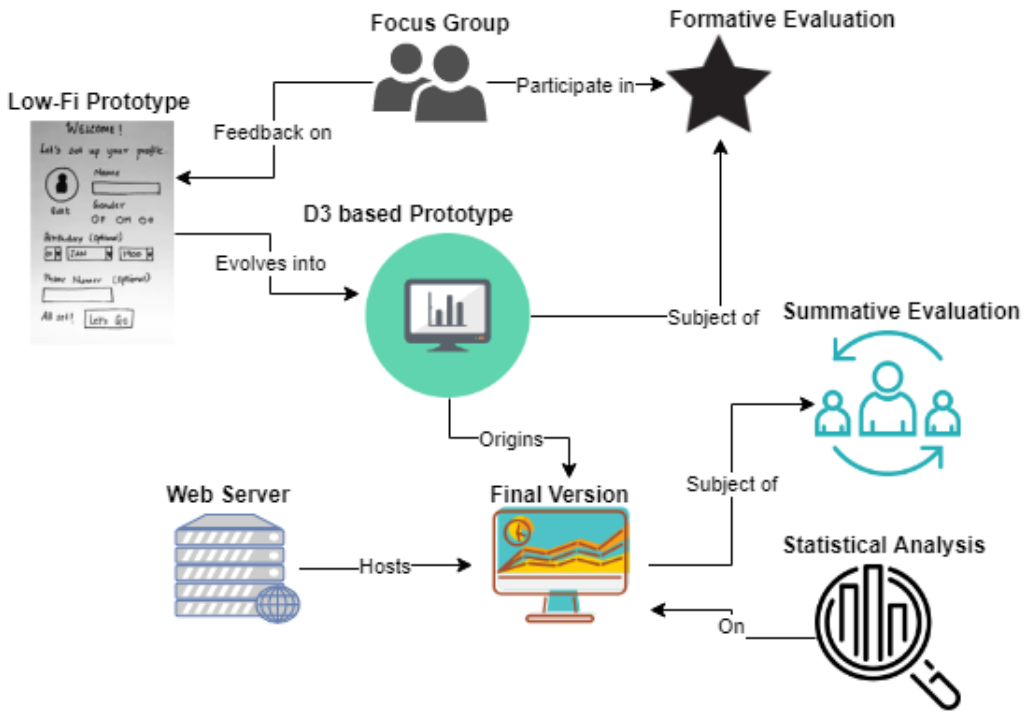

Figure 4.1: Conceptual Model describing reasoning and architecture of the solution.

The first conceptual idea developed of the prototype embodies the first mechanisms and interactions discussed in an early phase and obtained through the analysis of User Requirements. As shown in Fig. 4.1, a D3 based Visualization then emerges from the first conceptual idea and is subject of continuous Formative Evaluations, such as discussion sessions, brainstorming and Heuristic Evaluations, processes which provide important feedback to reiterate designing and improve the quality of the same Visualization. This implementation is based on JavaScript, HTML and CSS technologies, with the help of common libraries such as jQuery, Bootstrap and D3. This iterative process is based on the feedback and Usability problems found by the focus group until the design reach its final version, hosted on a web server for one more evaluation method. Thus, Summative Evaluation is then applied on the final version with target group users, where performance metrics are collected and further used for quantitative studies. The analysis of these metrics are useful for providing additional insight on the interaction between users and the design, as well as deriving conclusions and identifying further concerns to be addressed.

4.5 Metrics Studied

As introduced in Section 4.2, the feedback obtained through those processes led the developers to start conceptual modeling the solution, and one of first efforts made was to define which type of metrics and how those same metrics are going to be presented on the developed solution. From analyzing

the results from both Surveys and Interviews done with users from the target group in Section 4.2, the authors were able to synthesize and filter the ideas that would help users the most, while performing Usability related metrics analysis.

Therefore, a summary of the results was created and the list of metrics the authors feel that would provide the most insight while interacting with the visualization are the following:

- Completion Rate
- Completion Time
- Time spent on a determined area section
- Number of Errors
- Errors made on a determined area section
- SUS Rating

The items shown in the previous list represent the main metrics of which the infoVis will revolve around. Certainty, more derived information can be obtain from these particular ones, which will be further described and explained during Section 4.6 and Section 4.7.

Completion Rate is represented by a binary value, meaning if the user was able to complete the task handed (value of 1) or not (value of 0), number of errors, and along with errors made on a specific area, these three metrics cover up effectiveness assessment in the visualization. As for efficiency metrics, we decided to use both Completion Time, which translates as the time taken until the task is done, and Time spent on a determined area section, where this area could be a menu, form section or a comment one, for example. Lastly, for usability purposes and more specifically satisfaction, we defined our last metric as the SUS Rating, as it is one of the most common and efficient methods to measure the user's perception of the overall ease of the system [19], while being the most well known post questionnaire format, according to the Surveys handed in sub Section 4.2.1.

4.6 Early Prototypes

This Section explores the first procedures used in order to create the first conceptual idea and to define which ways catapulted the creation of such idea. The analysis, creation and mapping of the data that was used throughout the development of the solution as well as the initial sketches and mechanisms introduced are described bellow. Furthermore, the reasoning and feedback that triggered changes in the initial concept is shown, as well as the revised set of questions used as guidelines in the progress of the solution.

4.6.1 First Conceptual Idea

Before beginning with functional prototyping and starting with the design of the visualization itself, it is crucial that the nature of data presented is interpreted, from size to cardinality, as well as understanding the main goals and the audience that will interact with the solution. Besides, the cost associated (both time and money wise) with the development of a low-fidelity prototype is much smaller, compared to a functional prototype. Therefore, based on the feedback obtained through user requirements and on the initial set of questions that were derived from it, with the objective of obtaining feedback on some of the authors ideas for the visualization, some early ideas were designed through sketches.

One of the major requirements that we were able to withdraw from the discussion on the Chapter 3 was that it is important to maintain an equilibrium between the interactiveness and the clearness while providing information through an interface, meaning that the user should not be overwhelmed by the interactiveness presented and therefore stay focused on his objectives.

The initial reasoning behind the development of the prototype would include a tool for analysis comparison of different interfaces or applications. Although this idea would later be changed due to motives that are further explained in this section, and despite in this phase there were not any concerns regarding the layout of the visualization, the following figure Fig. 4.2 shows the initial sketch of what the authors foresaw in the appearance of the solution.

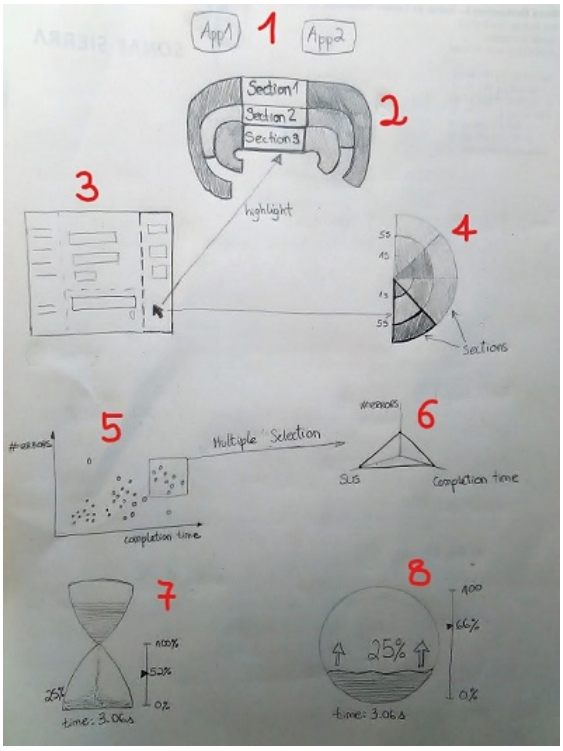

Figure 4.2: Initial sketch and Early Ideas.

As it can be shown through the presented figure Fig. 4.2, there is represented the initial view consisting of comparative Chart between two applications (Fig. 4.2, view 1) that measures not only the average time taken, but also the average number of errors in each section area, represented through bars' size and color (Fig. 4.2, view 2) of both applications. For further analysis, a selection is required.

After the selection of one interface/application, the information regarding the other one fades, and a new set of visualizations is shown. An overlay Visualization (Fig. 4.2, view 3) presents the structure of the interface through a snapshot, divided by section areas with a background color associated, related with the average time spent by users. By selecting or hovering over an area, other visualizations appear, highlighting its correspondent element. A Radial Chart (Fig. 4.2, view 4) can be used to analyze different groups of users based on the time they spent on a specific area, and therefore help answering questions such as "Do users make more errors when trying to spend less time on areas?", per example.

A Scatter Plot Visualization (Fig. 4.2, view 5) can help users explore patterns between all the entries, as well as identifying clusters, represented by groups that share identical features (in this case, number of errors and completion time) and comparing their SUS rating, represented by color on each dot. Furthermore, outliers are easily shown in this Visualization, and by selecting multiple dots, a Radar Chart (Fig. 4.2, view 6) is enabled, presenting the correlation between the three major metrics evaluated.

As mentioned throughout the document, the authors intended to make use of derived and statistical measures, and users interviewed and surveyed on Section 4.2.1 considered them of significant importance while exploring visualizations, while also providing useful overall information about the status and quality of each element analyzed. Consequently, we present two different approaches that can be used to answer questions such as "How much does it take until *2/3* of the users complete a certain task?", by selecting a threshold on the total percentage of the population, and when the visualization hits the cap, the time taken is presented as output. The two options considered and proposed consisted of a Hourglass Visualization (Fig. 4.2, view 7) and a Liquid Fill Gauge (Fig. 4.2, view 8). There were some concerns about the overall utility of these particular options, as well as the ability to transmit information without compromising the users focus, and therefore, as any initial conceptual idea, the design went through a series of discussions and brainstorming with the help of the focus group. These procedures would later result in changes for the first non functional prototype, as well as some consequent alterations in the set of questions, which are going to be further described in the next Section.

4.6.2 Changes and Revised Set of Questions

There were some modifications regarding the status of the design of the prototype, and these paragraphs describe why those steps and decisions were made and how they affected the design itself. The focus group, already mentioned, was composed by six elements who had worked and/or studied Usability and UX related areas, as well as background knowledge about Usability and about the needs

and objectives of this study, contributing with a stable source of feedback regarding the many phases of design.

After some informal sessions of discussion and brainstorming with members of the focus group, the authors decided that some ideas proposed were not necessarily converging with the main objective of this thesis, making them dispensable and thus discarded from the initial idea. Consequently, these changes originated a revision on the set of Questions previously defined.

Although the idea of comparing different interfaces or applications appeared interesting to some of the users interviewed on Section 4.2, during discussions sessions had with members of the focus group, the authors reached the conclusion that this functionality would increase the complexity of the visualization itself and might make its objective less clear. Since the main goal of this thesis diverges from a comparative analysis of interfaces, the development of a visualization with this functionality was set on hold, with the possibility of being brought back for future work, depending on users' feedback.

In the same way, the members of the focus group felt difficulty in understanding the purpose and the utility of the visualizations shown in Fig. 4.2, view 7 and view 8. The insight provided by these visualizations, which was to identify and quantify groups of tests which were under a specific value threshold, was seen as redundant, while using the argument that such information could be obtained with the help of simpler techniques such as filters, while not overloading the user visually. The changes mentioned in these paragraphs were also responsible for adjustments on the Set of Questions the authors pretended to answer through the help of the infoVis solution. Thus, following is described the revised set of twelve questions, which would serve as guidelines throughout the development phase.

- What is the time distribution between all sections for a given task?
- Is the section that takes users the most time also the most error prone?
- Do users make more errors when trying to spend less time on a given section?
- Are areas in the interface with higher cursor movement associated with more errors?
- Is there any individual test which is an outlier in terms of time/#errors?
- Which individual tests failed to complete their tasks?
- Are the most time consuming section areas associated with more errors?
- Where is the exact location of the errors made by an individual test?
- What is the most common error made across all tests for a given task?
- Does SUS rating correlate with time/#errors?
- What is the average time spent/#Errors/SUS Rating for a given group of individual tests?

- Do tests with similar metrics (time and #errors) share similar SUS Ratings?

The authors felt that this set of Questions assembled not only important conclusions regarding the Usability status of a given interface, but also kept a cohesive compilation of insight users pretend to obtain while analyzing such visualization.

4.7 Development phase

This Section describes how the development phase of the solution was carried out, including how the solution changed over time and what encouraged those changes. From the first low-fidelity prototype in paper to the last stage of the functional prototype as an interactive visualization, data mapping choices and interface layout changes, interactivity choices between idioms and Heuristic evaluations on the given prototype are some of the topics addressed in the following paragraphs.

The evolution of the presented solution was, as mentioned before, User feedback driven, and therefore each proposal and adjustment on the prototype, whether it was data mapping or functionality driven, was subject of formative evaluations by users inside the target group.

4.7.1 First Heuristic Evaluation

The first prototype was designed in paper, because it represents a quick and efficient way cost-wise to translate high-level design concepts into tangible and easily testable artifacts [2, 3]. Since this first prototype had the main goal of examining the visual representation and key elements present in content, the limited interactivity associated with this technique suited our purposes. With the analysis of a prototype in such early stages, both the authors and the testing participants involved could have a much clearer expectation about the upcoming stages of the solution, although the uncertainty during testing might hinder users from acknowledging what mechanisms are supposed to work or not.

With these thoughts in mind, the authors began to design the first low-fidelity prototype divided into three different visualization idioms, as shown in the following figure Fig. 4.3. The first idiom (Fig. 4.3, view 1) is represented by a Radial Bar chart, where the different user tasks currently analyzed are being displayed. As for data mapping, color is used to represent the number of errors, while length shows the average time taken to perform such task and the average SUS rating is shown with a secondary bar on top of the first one. After the user selects a given task, information regarding area sections are shown, as well as individual testing values, through an updateable Sunburst-like visualization. Again, size and color are used to map time and error measurements, both average (from a section perspective) and absolute values (for individual tests).

In order to further analyze user testing measures, and more specifically the several area sections used, a snapshot of a given interface carries information of user's cursor movement through color, as

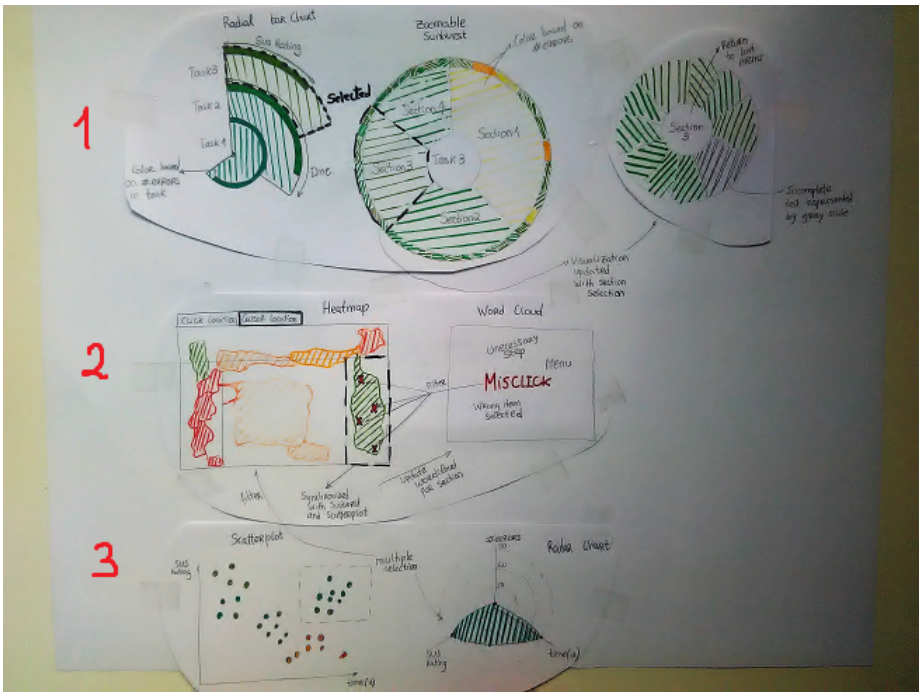

Figure 4.3: First low-fidelity prototype.

well as error locations by using error sign (red X's) (Fig. 4.3, view 2). An additional option can be set to alternatively track cursor movement of click usage. The authors believed an Heatmap visualization would accomplish the best metaphor for such data, which could easily help users identify areas associated with more errors and cognitive workload, and therefore areas of potential improvement of a given interface. To complement this idiom, an Index visualization shows the most common errors made by each user test, highlighted by font size, and that by selection would serve as a filter to the Heatmap, presenting only errors of the selected type.

The final idiom (Fig. 4.3, view 3) is destined to treat individual testing information, and therefore the authors decided to present it with the support of a Scatter Plot, which allows users to correlate the three main metrics involved (Time, Errors and SUS rating), to explore test results and identifying specific clusters or outliers. Additionally, a Radar Chart is used as guidance to further inspect a given group of tests, by representing their calculated mean values.

The low-fidelity prototype shown in Fig. 4.3 was then subject of a formative evaluation, and more specifically, the authors asked five Usability experts to perform an Heuristic evaluation based on Forsell's [15] set of infoVis Heuristics, shown back in Section 2.6. The experts were asked to highlight any particular problems found, along with a reference to the violated Usability principles and the severity of the problem itself, by using Nielsen's four step scale [16] (from 0 = I don't agree that this is a usability problem at all, to 4 = Usability catastrophe: imperative to fix this before product can be released). The were five evaluations, as it represents a suitable number of insight providers, while not compromising

neither time nor resources.

The evaluators were the asked to assess the given prototype and describe each usability problem found on it. This step was considered influential on the development of the proposed solution, since it help the authors building a better first functional prototype that focuses more on interactivity levels between every visualization on the dashboard. Several problems were reported, and thus, following are described the most important ones, by being the most commonly expressed or the most severe ones:

- Information Coding - Overall, there was a difficulty in interpreting the meaning for each of the colorings used. As an example, the evaluator mentioned: ”How many errors does light green represent?”. The solution proposed was to implement chart legends to each visualization and, of course, to keep them consistent throughout the dashboard.

- Information Coding - Specifically in the Radial Bar chart, the bars measuring each task's SUS and time might not be perceptible in terms of what bar belongs to each task. As a possible solution, axis implementations to ease information extraction and reconsider if using a second bar for SUS is the best option.

- Information Coding - Mentioned in the Sunburst, since the color is used to mapping number of errors, and since it is associated with a negative connotation, green colors might not be the best option, justified by the evaluator as ”a color more perceptible as a success one”. As a possible solution, it was mentioned to revised the color schemes and use red tones.

- Orientation and help - Regarding the Sunburst, it was mentioned it might not be intuitive for the user to regress to the main screen of the visualization when one decides to explore into detail by selecting a section. As a possible solution, a back button or a label should be used to remind the user he can move backwards.

- Orientation and help - On the radar chart, the scale used might not be ideal to represent three different metrics, as each as a distinct range of values. As for possible solutions, the evaluator mentioned the possibility of using percentages, or three scales.

- Consistency - In the Heatmap visualization, the Color scheme ranges from green to red, where the most usual and perceptive to users is to use cold colors for cold zones (less cursor action). For the solution, there should be a color scheme ranging from blue (colder) to red (warmer).

The evaluators provided not only feedback through the report of Usability problems, but were also interested in providing suggestions, not only for possible future functionalities, but also for the layout management. After each evaluation was finished, the evaluators mentioned important features and suggestions concerning the visualizations and the information those transmit: the consideration for temporal filters as it might improve exploration techniques; the careful usage of the space across the page, as well

as grouping visualizations with related purposes; and tracking current state were some of the insights provided.

4.7.2 Adjustments and Second Heuristic Evaluation

The results obtained from the first Heuristic evaluation encouraged the authors to continue the development of the prototype and, this time, to create a functional version with all the previous usability problems taken in consideration.

Regarding the dataset used, as mentioned in the previous Section 4.3, since the veracity of the data used would not contribute for the main purpose of this study, and in order to have a more UCD focused procedure to the given problem, the authors decided to apply a POC approach and therefore, to create a dataset suited for the visualization. The dataset was applied to fifteen individual Usability test results, and while not only covering times, numbers of errors and satisfactions levels, information about cursor usage was also used to derive other measures for visualization display. The number of fifteen instances of tests was considered suitable for the development of this solution and for the analysis of Usability testing results, as Usability expert Nielsen [11] proved existing a very weak correlation between the number of test users and Usability findings, along with the fact that most of the products used in Nielsen's study had a range of test users between five and fifteen.

The evolution of the first functional prototype started off with the development of the three different idioms present in the visualization, including basic interactions between each of them. Therefore, the following paragraphs summarize the evolution of each component of the visualization and the reasoning behind data mapping and functionalities applied.

The first idiom was composed by a Radial Chart and a Sunburst visualization, with the main purpose of selecting and beginning the exploration process for a specific task or area section that the user wants to assess. As shown in Fig. 4.4, the color scheme has been updated as suggested in the first Heuristic Evaluation for a red based color scheme, while satisfaction levels (SUS Rating) in the Radial Chart (left visualization) as been also changed, being represented through the width of each arc while, of course, complemented by respective tool tips. These changed were made based on the feedback received from users and were subject of further testing assessment. A "divide and conquer" approach was also used during this stage, relatively to the interactiveness implementations, and therefore a task selection on the Radar Chart correctly updates the information presented on the Sunburst, as well as individual interactions for each chart.

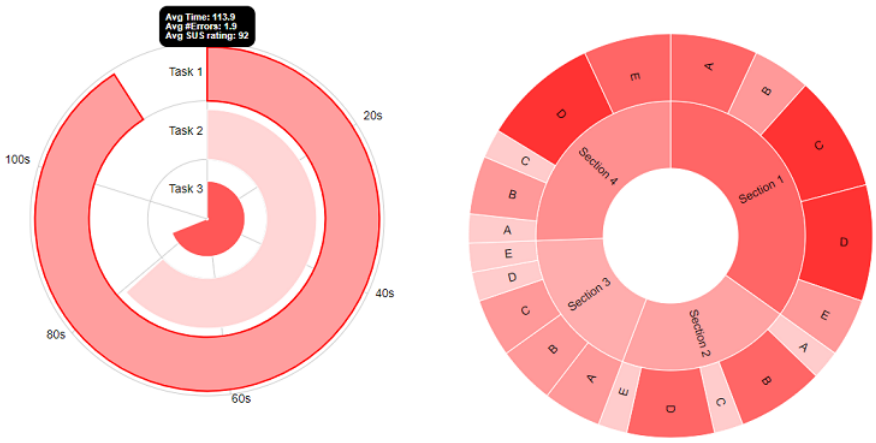

Figure 4.4: First idiom of the functional prototype.

As for the second idiom implementation, it was composed by a Heatmap visualization and complemented by an Index like visualization as a Word cloud. The objective of this idiom was to provide an intuitive way of identifying concerning section areas on the interface through mouse usage and error locations. These errors are defined by both a location and a label associated with each one, and therefore providing more insight for whoever is using the visualization, regarding what type of mistakes and where those are made. The Word cloud visualization has a complementary effect on this idiom, as it pinpoints to the user the most common errors committed, and furthering the exploration by providing a filtering tool for each error. The following Fig 4.5 presents the idiom described. Although in an early stage, this

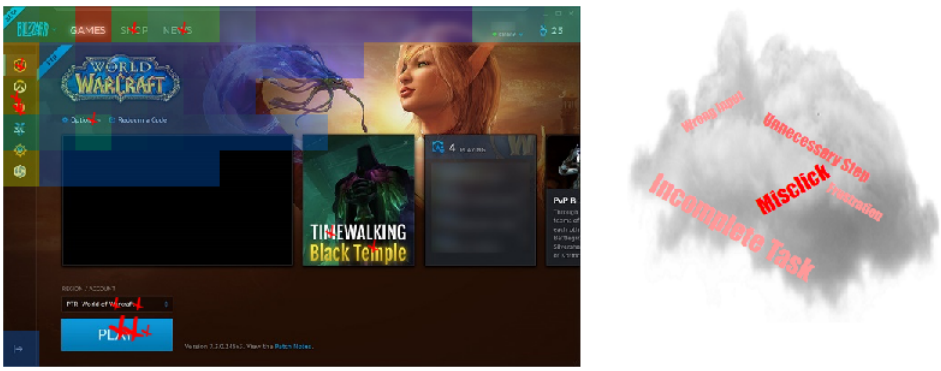

Figure 4.5: Second idiom of the functional prototype.

information can be contemplated in Fig. 4.5. As mentioned in the first Heuristic Evaluation, cold colors are represented with blue in the Heatmap, while error locations are shown through representative signs. As for the Word cloud, different font sizes are used to pinpoint different levels of frequency for errors. The design options adopted were once more, assessed by members of the focus group, even before the prototype reach the second Heuristic Evaluation, through discussions and observations.

Finally, the third idiom developed, and as firstly idealized in the first conceptual idea in 4.6.1, is

composed by a Scatter plot which visualizes and correlates the different Usability metrics assessed (Time, Errors and Satisfaction), along with a Radar Chart visualization that quantifies and compares the individual tests selected by the user in the Scatter plot. An early phase of the described idiom is shown in Fig. 4.6, as some functionalities and questions regarding information mapping were still in debate, such as the usage of still while representing individual tests in the Scatter plot, or the problem of normalizing the metrics with different ranging values in the Radar Chart. The development of the early functional

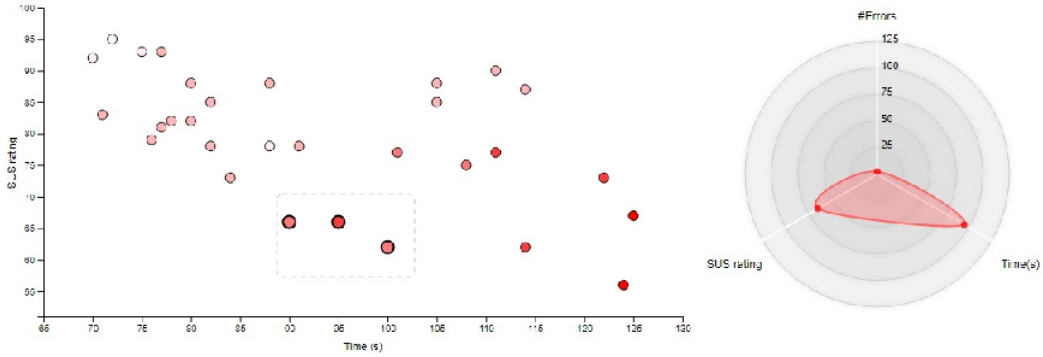

Figure 4.6: Third idiom of the functional prototype.

versions of the idioms shown in the last three figures (Fig. 4.4, Fig. 4.5 and Fig. 4.6) were focused on adding basic interactions between each one, while considering the potential of these versions against the low-fidelity prototype.

Such idioms were tested in different dispositions, bearing in mind the goals and objectives associated with each idiom described. Along with it, the focus group contributed throughout the development and evolution of the dashboard, as important questions regarding the design were discussed. One of those examples is the Radar Chart and how it would address the representation of three metrics with different connotations associated: while time and number of errors share a negative connotation, higher SUS Rating are generally associated with positive outcomes. As for the representation of individual tests that failed a given task, there was a reflection on using low saturation levels or gray scale to represent incompleteness - even though mixing colors would not be the best idea, the authors believed that with minimal learning it should not be a problem. Concerns regarding help for the user were also considered, as individualizing chart legends against sharing some, or using labels and buttons to describe the current state or to revert some actions.

As these questions were considered and debated between the authors and members of the focus group, a new version of the functional prototype would eventually rise, and a second Heuristic Evaluation would be used this time to not only assess the overall layout of the dashboard, but also how the data was managed and mapped, as well as the different interactive levels developed through the infoVis that matured.

As exhibited in the following figure Fig. 4.7, this new version of the functional prototype displayed a

different layout organized by functionalities: the user would be presented with a Radial Chart (Fig 4.7, view 1), and after the selection of a given task for analysis, the user would be then able to inspect and explore each of the following visualizations. Appropriate chart titles for contextualizing users, consistency on hover and selection mechanisms across the dashboard were adopted, color scales for all the visualizations were normalized, global chart legends and current status are some of the modifications that were made based on the previous insight given by the focus group and that needed validation through a new evaluation process.

Figure 4.7: Functional dashboard visualization used in the second Heuristic Evaluation.

As it is observable in Fig. 4.7 view 2, the current status is shown by displaying which task and section is currently under analysis, as well as a discrete chart legend based on red colors to describe the number of errors associated with each color, and therefore the gravity of each individual test performance. Regarding the representation of individual tests with incomplete tasks, the authors reached the conclusion that using low saturation levels of color and gray scales would not be perceptible for the users, particularly in situations were the number of errors associated with a given test result was zero, making it impossible to distinguish between a complete test and an incomplete one (both with white

color). Therefore, the usage of a dashed perimeter on for representing incompleteness on both the Sunburst visualization (Fig. 4.7, view 3) and Scatter plot (Fig. 4.7, view 4) was experimented and subject of evaluation with the experts. Regarding the Radar Chart (Fig. 4.7, view 5) visualization, a functionality to compare two different groups of individual tests was implemented, since users from the focus group showed interest in this characteristic and that it could further improve the quality of the information provided by the chart. Furthermore, three different axis were created, although the inversion of the metrics with negative connotation was necessary to provide the users with the natural feel of a Radar Chart - the larger the blob is represented, the better its property values are. Finally, the Word cloud (Fig. 4.7, view 7) visualization suffered some tweaks, primarily on font sizes and hover interactions to better differentiate the current error label selected, and on the angles for labels display - previous discussion sessions revealed that different angles for labels display, as perceived in Fig. 4.5, would confuse users and make it harder for them to acquire such information, which the authors believe can be mitigated by providing horizontal text.

The prototype previously shown in Fig. 4.7 was used for validation with five users contained in the target group. As minimum requirements, once again, the users were required to have previous experience with both infoVis and Usability fields, whether by attending college courses or working in UX related areas. Therefore, the authors consulted with five Usability experts and provided them a set of tasks in order to provide some context on the subject and to familiarize them with the given interface, while gradually introduce new mechanics and interactions between all the idioms presented. Additionally, the authors felt necessary to apply a *Think Aloud protocol*, considered an easy administrable, cheap and robust option [40, 41] to provide insight on how the users interact with the design or which elements are less perceptible to them. In such approach, the test participants were asked to use the system and complete the given tasks while constantly thinking out loud — by other words, to simply and continuously verbalize their thoughts as they interact with the design. In the end, a debriefing session was reserved to receive input from the participants on specific areas or elements on the interface that might need adjustments. The following paragraphs present the information obtained regarding the most common problems evidenced on the functional prototype in Fig. 4.7, as well as the heuristics violated, based once again on Forsell's [15] set of infoVis Heuristics, highlighted in Section 2.6.

- Information Coding - Chart titles could use a more contrasting color, as blue over black might be not the best choice. A more contrasting color, such as white, was suggested for the titles.

- Information Coding - Horizontal text facilitates the interaction with users and eases their information extraction. Not only for that reason, but also for a matter of consistency, users recommended using horizontal text on the Sunburst visualization (Fig. 4.7) for section labels.

- Flexibility - Overall users were happy with the amount of interactions present, but suggested fur-

ther optimizations, such as corresponding highlighted information between the Sunburst and the Scatter plot (Fig. 4.7, views 3 and 4), or filtering information on the Heatmap (Fig. 4.7, view 6) based on the given individual tests selected.

- Flexibility - In order to take into account different working strategies, habits and task requirements, users requested the option of zooming and exploring different sections by selecting them in the Heatmap (Fig. 4.7, view 6) visualization, as it provides users with more options to complete their tasks.

- Orientation and help - Having a global legend for all the charts was viewed as a weak choice as if forced users to recall information. Suggestions revolved around using individual chart legends and grouping visualizations and legends with similar and related purposes, such as the Sunburst and the Scatter plot (Fig. 4.7, views 3 and 4).

- Orientation and help - After selecting a different group of test users on the Scatter plot (Fig. 4.7, view 4), the Radar Chart (Fig. 4.7, view 5) shows two different blobs (as shown in Fig. 4.8), but some users reported difficulties in distinguish them, not only by the color choice, but documentation help was also missing. As suggestions, it was considered using different and more contrasting colors, as well as a chart legend with perhaps filtering mechanisms, which would further benefit its exploration.

- Orientation and help - A participant experienced difficulties in comparing two distant individual tests while exploring the Scatter plot (Fig. 4.7, view 4) visualization, and therefore suggested the implementation of guidelines, which would help users analyzing a give test by its metrics and to easily have an overview on how it compares with other individual tests.

- Spatial organization - While related to some of the other heuristic violations reported, the major consensus was that the space was poorly used, as there was a large blank space on the interface, which resulted in more scrolling down actions and making it difficult for users to interact with the design.

- Recognition rather than recall - Specifically on the Radial Chart (Fig. 4.7, view 1), a higher opacity on the current task selected might would help users to remember which state they selected, rather than having to recall their own actions.

- Recognition rather than recall - Again, in order to ease the whole interaction process between the users and the interface, information about the current state should be presented in the Navigation bar contrary to being presented between visualizations (Fig. 4.7, view 2). Also, the Navigation bar should be fixed on top of the page, providing a visible current state at all times to users.

- Prompting - During exploration specifically on the Sunburst visualization (Fig. 4.7, view 3), it might not be intuitive nor perceptible for users on how to retrieve backwards to the main page. For this particular problem, suggestions involved providing a label on the center of the visualization with a descriptive word of the action, such as ”MainPage”.

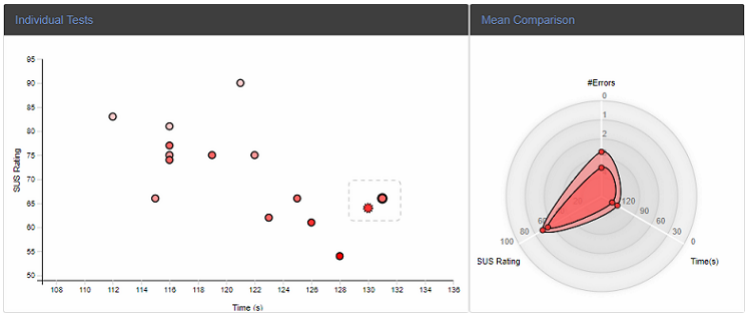

Figure 4.8: Second blob comparison after test group selected.

The second Heuristic evaluation provided a better insight on how users interact and envision the final version of the design, and resulted in important feedback on particular elements that seemed to hinder their experience. Those violations were taken in consideration and the authors proceeded to further improve the design and implement such changes, resulting in one last phase of the prototype that is further described in the next Section 4.8.

4.8 Final phase and User Testing preparations

After the second Heuristic evaluation was performed on the design, problems reported and improving suggestions were taken into consideration and implemented in the prototype. These changes, although resulting in the final phase of the design for the present study, the authors acknowledge that such work is never complete, particularly while adopting a UCD approach. On this section, the authors reserve some time to describe some extra decisions they made regarding the status of the design, as well as listing the final set of interactions implemented and the preparations for the final evaluation step.

One of the questions that had risen after second Heuristic evaluation report was that the color scheme used on the Radar Chart was not perceptible for the users, as they could not differentiate between the two blobs while comparing a group of users with the mean values of the whole population. Following this line of thought, the authors felt necessary to user two contrasting colors that were easily categorized by users, and thus not confusing their comparison. As discussed by *Colin Ware* [42], colors recommended for use in coding should be easily distinguishable even by color blind people and therefore have unique hues. From the set of twelve color represented in Fig. 4.9, the authors decided to use colors that would not be confused with any user in the rest of the color schemes, and thus decided to

use cyan and orange.

Secondly, there were some concerns regarding chart legends to represent number of errors, as a discrete chart legend was not suited to map decimal values, obtained whether through averages values or derived metrics, making it a reducer of the presented information. For this reason, the authors decided to use a continuous chart legend to represent the number of errors across all visualizations, due to consistency reasons and to being a more embracing solution.

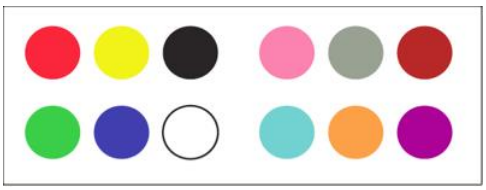

Figure 4.9: A set of twelve colors used in labeling [42].

The layout of the design suffered some changes, resulting from the feedback received in the second Heuristic evaluation. In order to remove unnecessary blank space and to reduce scrolling for the users, the final phase of the design and respective layout is presented in Fig. 4.10. The layout was adjusted according to the different objectives across each visualization: on the left side, it is possible to see visualizations that select and prompt other visualizations with exploration (Fig. 4.10, views 1 and 4); next to them are shown visualizations more exploratory related purposes, while keeping visualizations from the same idiom close: Heatmap with Word cloud (Fig. 4.10, views 2 and 3), and Scatter plot with Radar Chart (Fig. 4.10, views 5 and 6).

As perceived in Fig. 4.10, usability problems reported in the second Heuristic evaluation in Section. 4.7.2 resulted in changes based on the suggestions given by the users. Thus, the following paragraph will describe the major changes on each visualization, as well as interactions implemented with the rest of the design.

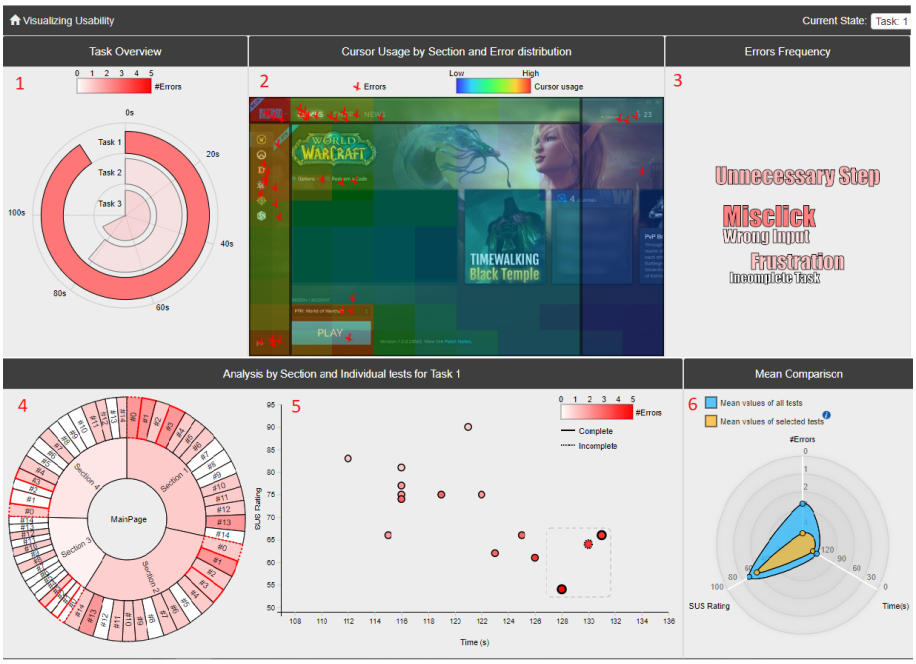

Figure 4.10: Final version of the functional prototype [1].

The Radial Chart (Fig. 4.10, view 1) uses is preceded by a continuous scale against the discrete used before, and has a more perceptible highlight after selecting a given task, by reducing opacity of the others. The Heatmap (Fig. 4.10, view 2) had some considerable changes: the addition of a corresponding legend on cursor usage levels and errors representation, as well as a hovering and selection of a given section, updating the rest of the interface accordingly to that same section. As for the Word cloud (Fig. 4.10, view 3), the font size and hovering interactions were consistently adjusted with the rest of the design. As mentioned in the previous Heuristic evaluation, the selection of a given individual test in the Sunburst (Fig. 4.10, view 4) visualization now affects the correspondent test in the Scatter plot (Fig. 4.10, view 5) and filters the errors presented in the Heatmap (Fig. 4.10, view 2). As a two way interaction, similar selections of individual tests on the Scatter plot (Fig. 4.10, view 5) provoke the same interactions across the interface. Finally, and as previously mentioned in this Section, the colors used were adjusted in the Radar Chart (Fig. 4.10, view 6), and a chart legend was added with a filtering mechanism, allowing users to highlight and suppress one or even both blobs.

With the final version of the prototype finished, the authors proceeded to the last part of the presented study, user validation. The description of the last evaluation step, as well as the results obtained and a discussion on the matter will be addressed in the following Chapter. 5

[1] The final version of the Visualization can be obtained in: https://fmorgad0.github.io/visualizing-usability/

5

Evaluation

Contents

This chapter intends to describe both evaluation methods utilized on the solution developed in Chapter 4, including results and insight that can be derived and concluded from this analysis. Firstly, the Usability tests conducted, as well as methodologies and protocols used are described throughout this Chapter, and are followed by and statistical analysis on the given results with conclusions and discussions on the insight obtained. Case studies were used in order to obtain further judgment on the status of the solution and each made is described and discussed regarding the feedback and suggestions given. The final Section is allocated for final conclusions and considerations concerning the solution created and how this evaluation process provides additional support to further development iterations.

5.1 Usability Tests

In order to assess the quality of the design achieved in Chapter 4 and to validate the solution presented, a quantitative study through a Summative Evaluation process is described in this Section. Since an UCD approach was used throughout this thesis, validating the solution near the target group allows the authors to assess its overall status and how far the solution is from being deployed and used daily.

Therefore, this Section describes the method used for evaluation, which was Usability Testing. From a characterization of the users to a description on the ambient and protocol used, the results of such tests and statistical analysis are presented along with the information and conclusions obtained. Finally, a discussion on the viability of the design and how further improvements can be made.

5.1.1 Description

Regarding the number of test users on Usability studies, Jakob Nielsen [11] defends that using five users almost always gets close to User Testing's maximum benefit-cost ratio. Since obtaining insight on the way the design interacts with users is not the only priority for this analysis, the authors decided to administer these Usability tests to a larger audience in order to obtain enough data to develop a quantitative study, and therefore to aim at statistical work. Naturally, feedback and suggestions were also retrieved from the test participants, aimed at collecting insights to drive the status of the solution, while being complemented by the information obtained from each Case Study.

For this study, the authors decided to test with at least twenty users with advanced experience with infoVis and the concept of Usability, while not having any previous experience with the design. The participants of these tests were all students or former students at *Instituto Superior Técnico* while also attending Usability and infoVis courses. The higher number of participants, contrary to the numbers used on previous formative evaluations (five on average), allowed the authors to get statistically significant numbers, and to obtain confidence intervals on the performance metrics, as well as reaching correlations and conclusions on the matter.

The lab used for the study had no background noise for participants to concentrate on the given tasks, while using a laptop provided by the authors. Further, a stopwatch was prepared as well as pen and paper for tracking users' difficulties and errors made throughout the test. Participants were briefly given a contextual introduction to the objective of this thesis and to the purpose of such Visualization, since they did not have any previous interaction with it, along with an informed consent for taking notes during the test (video nor audio were recorded). Shortly after, with the purpose of familiarizing participants with the Visualization's main functionalities, a period of five minutes was given for them to adapt and explore the design, resorting if needed to a set of three simple guideline tasks.

Following, the main phase began: the users were asked to read a set of tasks and to focus on completing each of them, one by one. Once the test participant decided to begin a given task, the facilitator asked him to complete it as efficiently as possible, while not being allowed to ask for help nor any questions. The authors decided to create a set of six tasks that involved every major interaction between the visualizations presented in the dashboard, as well as the main objectives regarding the extraction of information, while progressively adding a new layer of interactivity with each task. Following is listed the set of tasks presented to the Usability test participants:

- Identify the Task with the highest average time and the Task with the highest average number of errors.
- Identify the most frequent error made for Task 1. How many occurrences are there?
- Which individual test had the highest number of errors for Task 3, Section 3? How many errors does it have?
- Identify any incomplete individual test for Task 1. How many errors were made by said test across all sections? Which ones(label)?
- Select the two most time-consuming individual tests for Task 1, Section 1. What are their calculated mean values?
- Select every individual test with SUS Rating values ≥ 65 and ≤ 70 for Task 3, Section 3. How many errors of type "Misclick" where made in this group? Which Test ID made a "Wrong Input" error?

The protocol used was equal for all test participants and both completion time and errors were written down for each of the tasks completed. An error was considered whenever the participant made an incorrect interpretation of a visualization purpose, or whenever he took a wrong path towards the goal. There were a plurality of ways to achieve each task, and some participants managed to complete them in new manners, which were unacknowledged by the authors at the time. Naturally, these were not interpreted as errors, but as a flexible characteristic associated with the design.

After the users finished each of the six tasks, they were handed a SUS satisfaction questionnaire, a quick to administer and cheap tool to use for Usability measurement [20]. Consisting of a ten item questionnaire with five response options, from ”Strongly agree” to ”Strongly disagree”, this tool provides a valid, easy to administer and reliable tool to assess satisfaction, usability and learnability levels [43]. Therefore, this method was administered by the authors with the objective of evaluating the overall quality of the design based on the users' perception, and on different characteristics, such as complexity, consistency and learnability. Finally, users were asked to rate the perceived difficulty associated with each of the tasks completed, providing the authors with evidence for correlating results with the complexity levels.

After a debriefing session, the participants were free to ask questions regarding the state of the design, as well as providing suggestions and new functionalities they saw fitting in. Naturally, as perceived by the errors annotated during the test sessions, the authors discussed some of the problems the participants had, as well as possible solutions for those issues.

5.1.2 Results

As mentioned in the previous Section. 5.1.1, the authors considered as the main metrics for retrieval Effectiveness, Efficiency and Satisfaction. Thus, for each test, completion rates and number of errors were representative of Effectiveness, while completion time represented Efficiency and SUS score represented Satisfaction levels. The following paragraphs present the results obtained during the usability tests performed and statistical derived work, such as means, standard deviations, confidence intervals and correlations between said metrics.

Table 5.1: Confidence intervals on time taken to complete each task.

Task no.	Low(s)	High(s)	Margin of error(s)	Confidence level
1	8,683	11,846	1,582	95%
2	10,677	15,826	2,574	95%
3	18,970	27,319	4,175	95%
4	46,159	63,564	8,702	95%
5	47,163	56,541	4,689	95%
6	63,054	80,815	8,880	95%

Regarding the completion rate of the tasks administered to the test participants, only one user failed to complete Tasks 5 and 6, while the other nineteen users managed to successfully complete every task. As for the time taken by participants to complete each task, the previous Table 5.1 presents the results obtained. As perceived by the observation of the table and with a confidence level of 95%, Task 1 takes between 8,683s and 11,846s; Task 2 takes between 10,677s and 15,826s; Task 3 takes between 18,970s and 27,319s; Task 4 takes between 46,159s and 63,564s; Task 5 takes between 47,163s and 56,541s and Task 6 between 63,054s and 80,815s. From this table can also be derived that the time

needed to complete each task increases simultaneously with each task, with the exception of the fifth, which may be justified due to the disparity in the margins of error encountered between the forth and the fifth tasks.

Table 5.2: Confidence intervals on number of errors for each task.

Task no.	Low	High	Margin of error	Confidence level
1	0,042	0,458	0,208	95%
2	0,311	0,789	0,239	95%
3	0,080	0,520	0,220	95%
4	0,267	0,833	0,283	95%
5	0,080	0,520	0,220	95%
6	0,393	1,007	0,307	95%

Table 5.2 presents information regarding the number of errors made by the users for each of the tasks handed. As it can be observed on the presented table, and again with a confidence level of 95%, users commit between 0,042 and 0,458 errors in Task 1; between 0,311 and 0,789 errors in Task 2; between 0,080 and 0,520 errors in Task 3; between 0,267 and 0,833 errors in Task 4; between 0,080 and 0,520 errors in Task 5 and between 0,393 and 1,007 errors in Task 6. From this information we can derive that tasks 2, 4 and 6 are more error prone for the users, while not necessarily taking users the most time (from observation of Table 5.1). This situation might be justified by the complexity or difficulty while interacting with the specific visualizations associated with the given tasks.

Table 5.3: Confidence intervals on SUS score.

Low	High	Margin of error	Confidence level
69,528	77,722	4,097	95%

As mentioned in the beginning of this Chapter, Satisfaction levels were also subject of evaluation near the test participants. In order to assess Satisfaction, the authors used a SUS satisfaction questionnaire, with a score between 0 and 100. The authors stress the fact that, according to Brooke [20] and Sauro [43], such scores are not to be interpreted as percentages, but in fact as a percentile. With the average score being 68, a score lower than 51 places a product in the bottom 15% and a score higher than 80.3 in the top 10%. From the examination of Table 5.3, the design used in the Usability tests scores, with a confidence level of 95%, between 69,528 and 77,722. Thus, this result suggests that the design ranges above average score, with some design flaws and improvements needed but with an acceptable Satisfaction near the users.

In order to further retrieve additional information about the Usability test performances from the participants, as part of the post questionnaire used, the authors asked users to rate each task form a difficulty point of view. The objective was for the authors to perceive any particular disparity in the complexity of the tasks handed, as well as obtaining proof regarding the increasing level of difficulty with

Table 5.4: Confidence intervals on difficulty levels for users in each task, ranging from 1 (hard) to 5 (easy).

Task no.	Low	High	Margin of error	Confidence level
1	4,691	5,000	0,209	95%
2	4,433	4,967	0,267	95%
3	4,075	4,625	0,275	95%
4	3,016	3,784	0,384	95%
5	2,842	3,758	0,458	95%
6	2,144	2,856	0,356	95%

each task. With five different possible responses, users classified each task performed from 1 (hard) to 5 (easy). As shown in Table 5.4, and again using a confidence level of 95%, the tasks used in the Usability tests are considered progressively harder for users, with the exception once more of Task 5, which shows similar results to Task 4, while having a greater margin of error. Even though the tasks were presented in this order, the novelty of each one and the new interactions presented throughout the Usability test session displayed an increasing level of complexity for users to complete them. This information is somehow complemented by the increasing time users took to accomplish the same tasks (Table 5.1).

After gathering proof from the participants on how the difficulty levels of each task are consequently harder, while also obtaining information on time and number of errors users had, there were some further analysis and questions regarding the results the authors wanted to corroborate. As introduced in the beginning of this chapter, the authors intended to perform Usability tests with a large enough audience in order to provide enough information to present quantitative results, which is shown in the previous paragraphs. Besides that, there was a need to further assess the status of the design and to correlate some of the metrics that were obtained throughout the tests.

Thus, one of the questions that needed validation was verifying how the time taken would correlate to the number of errors. With 20 test participants, and with 6 different tasks to extract results, the authors decided to compute *Pearson's Correlation* on those 120 entries. Considered the most used method to determine the strength of the relationship between two variable, *Pearson's Correlation* [44] outputs a value between -1 and 1: while a coefficient value of 1 is representative of a strong positive relationship, a value of -1 represents a strong negative one. On the following Figure. 5.1, the results obtained are displayed.

		TIME	ERRORS
TIME	Pearson Correlation	1	,260**
	Sig. (2-tailed)		,004
	N	120	120
ERRORS	Pearson Correlation	,260**	1
	Sig. (2-tailed)	,004	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.1: Pearson's Correlation Coefficient value between time taken and number of errors for each task. Output obtained from IBM SPSS software [1].

As perceivable through the analysis of Fig. 5.1, there is in fact a positive correlation associated with the relationship between time taken and errors made, meaning it is observable a increasing number of errors while the users perform more time consuming tasks. The correlation is significant at the 0.01 level (2-tailed) and thus, with 99% confidence, the authors found a correlation of 0.260 between these two variables. Although a positive correlation was found, the small coefficient suggests a weak positive relationship (.20 to .29) between the time taken to complete a certain task and the number of errors made, leading to a weak dependency. The authors believe that such results may be influenced by the learning pattern of tests users while completing the first tasks, easing the process once reaching the latter ones.

As mentioned throughout the Chapter and more concretely on Section. 5.1.1, the Usability tests were used not only to quantify and perform statistical studies on the test participants' results, but served also as a way to identify flaws in the design, and thus every error made by users was categorized. Furthermore, after the test session was complete, users were encouraged to provide feedback on how their experience was, from identifying more complex areas in the design to suggestions on how to work on those matters. As a result of these mechanisms, the following list presents the most common Usability problems found in the Visualization Dashboard, whether by errors made or by feedback provided from the users.

- Graph legend gradient on #errors incompatible with hover interactions, since both use red based colors.
- Selection boxes used on the Scatter Plot Visualization should provide additional help regarding the range of the values selected.
- The Scatter Plot Visualization should further an option for filtering tests on specific characteristics, such as completeness.

[1] IBM SPSS statistics tool can be obtained in the following URL: https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software.

- Representing incomplete individual tests with a Dashed line perimeter presents a difficult data mapping option for users to identify such tests.
- The Word Cloud Visualization should provide a filtering functionality for a specific type of error.
- Regarding the Sunburst Visualization, a ”Section Index” label was suggested instead of ”Main Page” for backtrack button once a Area Section is selected, since it is more representative of the objective of the button.
- There are some scalability concerns regarding the usage of the Sunburst Visualization: what if more tests, or even area sections are to be explored?
- Although this information can be indirectly achieved through other metrics, there is not an explicit way of obtaining the total number of errors made for a specific task.
- The Word Cloud Visualization should present error labels ordered by number of occurrences.
- A specific Area Section selected should stay highlighted on the Heatmap Visualization, helping the user keeping track of the current state of exploration.
- Errors representation on the Heatmap Visualization might sometimes be confused with the background, specially on areas with high cursor usage (both red based colors).

The presented list of Usability problems is representative of the most common difficulties users had while performing the Usability tests, and takes in consideration as well some of the suggestions presented to the Authors regarding data mapping and overall functionalities.

5.1.3 Discussion

As mentioned in Chapter 1, the main objective of this thesis was to explore and study interactive ways of assessing Usability measures through Information Visualization, and therefore the design developed throughout Chapter 4 was not only subject of a series of formative evaluations near the target users, but needed also validation through a Summative Evaluation process. A quantitative study was important in a way that it enabled an analysis of the Usability of the solution developed by aiming at statistics. This way, this procedure involved twenty test participants that allowed the authors to derive conclusions based on the results obtained and on how those users interacted with the design, through Usability related metrics such Effectiveness, Efficiency and Satisfaction. While Usability expert Nielsen [11] suggests that there is not an ideal number of test participants, in order to perform quantitative studies and to gather enough data to retrieve statistically significant numbers, at least a number of 20 users should be used.

The Authors were able to obtain not only significant information regarding users performance over the design developed, as confidence intervals on Effectiveness, Efficiency and Satisfaction metrics, but also

to collect different conclusions regarding the interactions between users and the solution. As displayed in Section 5.1.2, the most error prone tasks were associated with the Usability problems reported, as well as those with higher margins of error, due to disparate values across test participants. As an UCD development driven approach is being used, such information that is collected is used to further reiterate development phases the feedback obtained from users is important in a way that complements and improves the overall Usability of the solution being developed.

There was a need to investigate on how particularly the time taken by users to complete the different tasks was related to the number of errors made. By using the Pearson's Correlation Coefficient, the Authors obtained a value of 0.260 (as shown in Fig. 5.1), which represents a weak uphill correlation between both variables. This sustains the Authors' belief that the low value obtained suggests that users are able to learn and adapt to the design fairly quickly without compromising their Effectiveness.

Furthermore, the Authors were able to gather a set of Usability problems regarding the current state of the design and that felt the most important for users. Such information allows further development through new iterations and versions with the expectation of a more usable solution.

5.2 Case Studies

As a method of complementing Usability Testing, the authors decided to conduct three Case Studies on the solution design. The case studies were administered to people with extended knowledge and experience on both Usability and infoVis areas, and were made with the objective of obtaining additional insight on the viability of the visualization for practical use, as well as some suggestions regarding future possible functionalities to add. As concluded in the previous Section 5.1, the design will continue to improve in order to provide a better experience to the target group, but the Case Study participants might think of new functionalities and utilities for the described design which will aid them collecting and interpreting Usability related data. As participants work on related areas, the authors intend to substantiate the value of the developed solution for the target group and which purpose it serves regarding the daily tasks performed on the analysis of Usability metrics. Additionally, it is important to collect feedback on how such tool can be further complemented, in order to maximize utility.

5.2.1 Description of each study

During the Summative evaluation phase, the authors decided to obtain additional information and feedback from people that could use a tool of this nature, as well as insight on how they perform and interact with it. Three different case studies were performed with people working on educational areas, from teaching to research, and following is described how those sessions took place, as well as insight gathered from each one.

For the actual approach used, the authors presented the participants with the design and provided them with a training period of five minutes, in order to adjust and adapt themselves to the existing layout and to the functionalities associated with each area of the Dashboard. After, a set of tasks similar to the ones presented in Section 5.1.1 served the purpose of guidelines to introduce every interaction and functionality between every Visualization shown. Although the procedure adopted was similar to the one used for Usability Testing, the participants of the Case Studies were not subject of evaluation regarding the collection of performance metrics. Since the main goal of these studies were to collect additional insight on the utility of the solution developed for the target group, its purpose diverged from the quantitative study developed in the Usability Testing. In the end, participants were given additional time to explore any particular area of the design, and could express their current feelings regarding their interaction with the Dashboard, whether by identifying shortcomings or make some remarks on its usage.

The first Case Study was performed with a Professor and Researcher associated with the areas of Computer Graphics (CG) and HCI, with extended knowledge on user interface and Usability, and thus the authors felt that feedback from a person with such background could be meaningful for this study. From the observation during the training and tasks resolution, the participant felt difficulty while interacting with some particular visualizations and extracting information from those, as he mentioned later in the discussion phase. The participant felt that the Dashboard was well integrated and all visualizations interacted accordingly, ”(...) the components interact well with each other and we can notice the integration between all of them”, but struggled and was sometimes overwhelmed with the amount of information presented and with some visualizations in particular, ”(...) there is too much information represented at the same time.” and ”The Word Cloud sometimes becomes confusing.”. Besides, the participant mentioned that he is not sure about the color gradient for representing the number of errors, as the contrast between the different values were sometimes not perceptible for him immediately. As for the utility and purpose of the solution, he showed interest in a tool that would help to identify troublesome areas and determine the resolution approach through the analysis of the errors associated with such areas, and as a recommendation, added that ”It would be pleasant for future work to implement a mechanism to prioritize a given error or suggest areas to concern ourselves first”.

The second Case Study was taken with a Researcher who is starting is career on the fields of HCI and had taken specialization courses related with both infoVis and Usability, who had an interest on giving feedback and recommendations on the solution developed. The participant demonstrated a different level of accommodation in comparison with the previous Case Study participant, and felt confident with the majority of the interactions within the design once he had trained with it. These aspects were corroborated with his remarks on the solution, as he mentioned ”Once I got used to it, the experience felt much smoother.”. Regarding the utility and general thoughts on the design he added ”It

focuses on very important metrics, (...) I feel that this is a very interesting contribution.". The participant shown interest and asked about the readiness of the solution to be used for testing his own products, and during this discussion he showed some concerns as there "(...) might exist some difficulties generalizing the work done". Thus, besides the adjustments needed regarding Usability problems, the participant expressed the need for future work to improve the overall abstraction and generalization of the tool, and thus simplifying the distribution and the analysis of different products.

The last Case Study had a similar approach to the previous ones described, as the authors talked to a Researcher and Assistant Professor who lectures courses on HCI and infoVis related areas. As the second Case Study developed, this participant showed also some dexterity regarding the resolution of the tasks presented and the adaptation process to the design, justifying that his performance was somehow biased due to previous experiences, mentioning "I've dealt with the majority of these visualizations in the last semester as I lectured an infoVis course.". His main challenge resided with discovering how these visualizations interacted with each other, and how to explore and interpret some information displayed. The participant felt that the various functions in the system were well integrated and that the major flaws were associated with data representation, as "(...) the color scheme might cause some confusion with the users, particularly the errors quantity gradient." and that spotting incomplete individual tests with resort to a dashed perimeter "(...) could be somehow challenging." and that "(...) such problem could be mitigated with the usage of filtering options regarding the inclusion of incomplete tests on the display". Finally, a discussion on the importance of a tool with these purposes was held, and the participant pinpointed that it could "(...) have an interesting contribution on an academic context (...)." by using it on HCI courses on the analysis of Usability test results"

5.2.2 Discussion

The Case Studies performed served as a complementary method of evaluating the solution developed, while also allowing the authors to collect insights on how such tools could be used by the future users in their daily tasks. From the results obtained, the authors could infer some conclusions regarding the state of the solution and what future steps should be taken in order to continuously improve and to fulfill the target group's requirements.

Although the participants involved in the Case Studies had positive remarks regarding the integration of the various functions across the Dashboard, there were some concerns with a few visualizations and some choices made for the mapping and representation of data. The Word Cloud was particularly pinpointed from one participant as being sometimes confusing and potentially hard to differentiate between the different levels of error frequencies, and although the usage of color assisted the issue in hand, the color scale used could be further improved. The color gradient scale was not only mentioned in the previous problem, but throughout all Case Studies, and thus the authors acknowledge that as being one

of the most important priorities for further work. Furthermore, one Case Study revealed a difficulty in the identification incomplete tests in the design. The problems reported in each process, whether by Usability Testing or by Case Studies, revealed some similarities and matching concerns, and for that reason both approaches presented themselves useful for the authors regarding the assessment of the solution developed.

These sessions were also important as a matter of identifying potential fields of interest for the usage of the presented solution. A Dashboard Visualization with the effort of improving the analysis of Usability Testing related metrics was viewed as a promising tool not only for educational purposes but also to aid users to assess the Usability of solutions developed in Research contexts. Additionally, the potential shown by the solution allowed participants to provide new possible features to be included, such as suggestion based functionalities by pinpointing the most concerning areas and important types of errors.

5.3 Final considerations

Throughout this Chapter, the authors described and outlined the evaluation processes held with the evaluation phase of the solution developed in Chapter 4. Firstly, and in order to obtain statistically significant numbers and tight confidence intervals, a series of twenty Usability tests served as subject of a statistical analysis, with all participants having experience on both infoVis and Usability areas. The second evaluation method used consisted of three Case Studies performed without any kind of metrics retrieval, while focusing more on the insight and suggestions given by the participants, whether it was regarding the layout of the Dashboard, or particular interactions between the differing idioms.

By analyzing each evaluation method performed and the results obtained from both, the authors can derive and sustain some insight and considerations regarding the prototype. As perceivable through the observation of Table 5.1 and Table 5.2, the fourth task had disparate results across the population used, represented by high Margins of error, particularly while comparing with tasks that participants rated as more complex. The authors conclude that high intervals on both time and number of errors associated with the fourth task are justifiable by the results obtained from participants that struggled particularly with the color mapping and identifying incomplete individual tests, information which can be corroborated by the same concerns mentioned by some of the Case Study participants. The resemblances shown on the difficulties participants had on both evaluation methods not only confirm but also highlight the Usability problems which need the fastest intervention.

Furthermore, and following the same lines of reasoning, the Satisfaction levels (Table 5.3) across all Usability test participants revealed that, after normalizing such values and converting them to a percentile rank, and according to a Sauro's graph that relates 500 studies' percentiles with the SUS

scores [43], the presented design ranks above average. The positive reaction regarding the solution's Usability is also mentioned on the Case Studies performed, particularly after users had adapted to the design. This insight, along with the Satisfaction levels obtained, suggest that, although the design is perceivably usable for both Usability Testing and Case Study participants, still holds some Usability flaws and needs further improvements in order to provide a suitable experience for users and allowing them to complete their daily tasks.

The authors collected information on the perceived difficulty for participants associated with each task (as shown in Table 5.4), and the results opened an hypothesis that was later confirmed: "Would an increase in time associated with a given task cause an escalation for the number of errors?". The authors computed the *Pearson's Correlation* coefficient between these two variable and managed to prove that, although a small positive correlation exists, a value of 0.260 obtained (refer to Fig. 5.1) represents a weak dependency between both, and therefore the authors believe that such value is indicative of a learnability factor associated with the design, allowing users to adapt and improve their performance. Thus, on both evaluation approach used throughout this Chapter, the results substantiate that users will not necessarily make more errors on complex tasks that consume more time, as they adapt themselves to the design with some training.

As mentioned in the previous paragraphs, the three instances of Case Studies performed provided a source of insight and concerns regarding the overall Usability of the design that complemented the results collected from the User Testing. Furthermore, participants contributed with suggestions and concerns on the design, from changes that should improve UX and new functionalities to the utility of such tool flexibility goals, matters which will be further discussed on the following Chapter.

6

Conclusions and Future Work

Contents

This work presented a Visualization solution with the purpose of studying and assessing Usability measures through Information Visualization techniques, with the belief that infoVis applied on Usability can provide a different paradigm in the analysis of a product's quality during its development stage, and as mentioned by Munzner [1], infoVis allows users to analyze data when they do not know exactly what questions they need to ask in advance, and therefore derive information that would be difficult otherwise. The authors believe infoVis can be applied in UCD approaches and be a reliable tool in the analysis of Usability related metrics obtained from tests with users, reducing cognitive workload on users while extracting information from data with such complex and subjective nature.

The main objective of this thesis was considered achieved, as Information Visualization was applied in the analysis of Usability related metrics and the Visualization Dashboard developed was validated by target group users, who rated the solution with an above average Satisfaction, as shown in Chapter 5.

The results obtained from Usability Testing supported additional statistical studies. By revealing a weak dependency between the time taken to complete a given task and the number of errors made, the authors found that although more complex tasks consumed more time for users, those were not necessarily more error prone for them, which leads to the conclusion that the design provides a stable learning curve for new users. Case Studies represented a significant source of conclusions and insights regarding the state and utility of the presented solution. Participants were chosen based on their experience with both infoVis and Usability areas, and acknowledged the importance and utility of having such a tool to analyze Usability data, which some suggested being sometimes a complex task. Furthermore, they recognize the importance of the metrics that are being visualized throughout the Dashboard, although some concerns were shown regarding the overload of information shown at a time, which deviates from the main idea of using infoVis in the first place.

A plurality of utility areas and practical usages for the solution were suggested, and such tool was acknowledged by Case Study participants to be particularly useful in an educational context to help students identify flawed areas in their designs, as well as representing a supportive tool for researchers on the interpretation of Usability Testing results collected. Such areas were not initially thought of, and such testaments enhance, in the authors eyes, the contribution of this study.

6.1 Future Work

The work presented leaves room for both improvements and new features to further improve the quality of the solution. Results obtained from Usability Testing revealed some concerns with some of the visualizations regarding their scalability, particularly the Sunburst one, which may hinder the analysis for users while performing exploratory tasks. Additionally, other flaws reported by users expose flaws concerning data mapping, as the color scheme and representation for some metrics (number of errors

and task completion). These problems contributed to deter results for some users, who struggled to complete some tasks, and in a particular case, even resulting in failure. Furthermore, enhancements and fixes based on the Usability problems found and reported by users should be implemented, having in mind the goal to provide a better UX for the target audience. Besides, the scalability problems associated with some visualizations reside has an important step towards the flexibility of this tool.

The group of visualization techniques utilized on the resulting Dashboard solution were subject of validation, and the quantitative study performed would benefit from further refinements and statistical analysis, such as comparing results with two different populations: one from the focus group and other from outside of it. Additional studies could be done on future iterations based on the Usability problems found on the present version of the solution, with the example of a comparative analysis regarding the performance of two or more design choices.

Finally, future work on this study includes the distribution of such tool, which helps users assessing the Usability of their design on different areas, such as educational and research, as well as testing with users on different contexts, while using real data.

Bibliography

[1] T. Munzner, *Visualization Analysis and Design*, ser. AK Peters Visualization Series. CRC Press, 2015. [Online]. Available: https://books.google.de/books?id=NfkYCwAAQBAJ

[2] J. Preece, Y. Rogers, and H. Sharp, *Beyond Interaction Design: Beyond Human-Computer Interaction*. New York, NY, USA: John Wiley & Sons, Inc., 2001.

[3] A. Dix, J. Finlay, G. Abowd, and R. Beale, *Human-computer Interaction*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.

[4] ISO, "Iso 9241-11:1998(en), last visited 14/05/2015," 1998, uRL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en.

[5] J. Nielsen, "Parallel & iterative design + competitive testing = high usability," January 2011, uRL: https://www.nngroup.com/articles/parallel-and-iterative-design [Retrieved March, 2017]. [Online]. Available: https://www.nngroup.com/articles/parallel-and-iterative-design/

[6] ——, "User satisfaction vs. performance metrics," October 2012, uRL: https://www.nngroup.com/articles/satisfaction-vs-performance-metrics [Retrieved March, 2017]. [Online]. Available: https://www.nngroup.com/articles/satisfaction-vs-performance-metrics/

[7] J.-Y. Mao, K. Vredenburg, P. W. Smith, and T. Carey, "The state of user-centered design practice," *Commun. ACM*, vol. 48, no. 3, pp. 105–109, Mar. 2005. [Online]. Available: http://doi.acm.org/10.1145/1047671.1047677

[8] S. Wilson, M. Bekker, H. Johnson, and P. Johnson, *Costs and Benefits of User Involvement in Design: Practitioners' Views*. London: Springer London, 1996, pp. 221–240. [Online]. Available: http://dx.doi.org/10.1007/978-1-4471-3588-3_15

[9] S. Wilson, M. Bekker, P. Johnson, and H. Johnson, "Helping and hindering user involvement — a tale of everyday design," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '97. New York, NY, USA: ACM, 1997, pp. 178–185. [Online]. Available: http://doi.acm.org/10.1145/258549.258699

[10] J. D. Gould and C. Lewis, “Designing for usability: Key principles and what designers think,” *Commun. ACM*, vol. 28, no. 3, pp. 300–311, Mar. 1985. [Online]. Available: http://doi.acm.org/10.1145/3166.3170

[11] J. Nielsen, “How many test users in a usability study,” June 2012, uRL: https://www.nngroup.com/articles/how-many-test-users/ [Retrieved February, 2018]. [Online]. Available: https://www.nngroup.com/articles/how-many-test-users/

[12] J. Nielsen and R. Molich, “Heuristic evaluation of user interfaces,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’90. New York, NY, USA: ACM, 1990, pp. 249–256. [Online]. Available: http://doi.acm.org/10.1145/97243.97281

[13] J. Nielsen, “Enhancing the explanatory power of usability heuristics,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’94. New York, NY, USA: ACM, 1994, pp. 152–158. [Online]. Available: http://doi.acm.org/10.1145/191666.191729

[14] J. Nielsen, “Usability inspection methods,” J. Nielsen and R. L. Mack, Eds. New York, NY, USA: John Wiley & Sons, Inc., 1994, ch. Heuristic Evaluation, pp. 25–62. [Online]. Available: http://dl.acm.org/citation.cfm?id=189200.189209

[15] C. Forsell and J. Johansson, “An heuristic set for evaluation in information visualization,” in *Proceedings of the International Conference on Advanced Visual Interfaces*, ser. AVI ’10. New York, NY, USA: ACM, 2010, pp. 199–206. [Online]. Available: http://doi.acm.org/10.1145/1842993.1843029

[16] J. Nielsen, “Severity ratings for usability problems,” January 1995, uRL: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [Retrieved February, 2018]. [Online]. Available: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

[17] ISO, “Software engineering - product quality part 4, ISO/IEC 9126-4,” International Organization for Standardization, Tech. Rep., 2001.

[18] J. Sauro, “Performance satisfaction and perception satisfaction,” February 2010, uRL: https://measuringu.com/test-task-sat [Retrieved March, 2017]. [Online]. Available: https://measuringu.com/test-task-sat/

[19] J. Sauro and J. R. Lewis, “Correlations among prototypical usability metrics: Evidence for the construct of usability,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’09. New York, NY, USA: ACM, 2009, pp. 1609–1618. [Online]. Available: http://doi.acm.org/10.1145/1518701.1518947

[20] J. Brooke, *"SUS-A quick and dirty usability scale." Usability evaluation in industry*. CRC Press, June 1996, iSBN: 9780748404605. [Online]. Available: https://www.crcpress.com/product/isbn/9780748404605

[21] T. Dutta, "Evaluation of the kinect™ sensor for 3-d kinematic measurement in the workplace," *Applied Ergonomics*, vol. 43, no. 4, pp. 645–649, 2012.

[22] P. Carayon, T. Wetterneck, A. Rivera-Rodriguez, A. Hundt, P. Hoonakker, R. Holden, and A. Gurses, "Human factors systems approach to healthcare quality and patient safety," *Applied Ergonomics*, vol. 45, no. 1, pp. 14–25, 1 2014.

[23] M. Riveiro and G. Falkman, "Evaluating the usability of visualizations of normal behavioral models for analytical reasoning," in *Proceedings of the 2010 Seventh International Conference on Computer Graphics, Imaging and Visualization*, ser. CGIV '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 179–185. [Online]. Available: http://dx.doi.org/10.1109/CGIV.2010.34

[24] J. W. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.

[25] A. Gelman and A. Unwin, "Infovis and statistical graphics: Different goals, different looks." [Online]. Available: http://www.stat.columbia.edu/~gelman/research/published/vis14.pdf

[26] A. Tervakari, K. Silius, J. Koro, J. Paukkeri, and O. Pirttilä, *Usefulness of Information Visualizations Based on Educational Data*, ser. IEEE Global Engineering Education Conference. IEEE, 2014, pp. 142–151, contribution: organisation=mat,FACT1=1¡br/¿Portfolio EDEND: 2014-09-30¡br/¿Publisher name: IEEE.

[27] M. H. Falakmasir, I.-H. Hsiao, L. Mazzola, N. Grant, and P. Brusilovsky, "The impact of social performance visualization on students," in *Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on*. IEEE, Jul. 2012, pp. 565–569. [Online]. Available: http://dx.doi.org/10.1109/icalt.2012.218

[28] Y. Chen, Q. Chen, M. Zhao, S. Boyer, K. Veeramachaneni, and H. Qu, "Dropoutseer: Visualizing learning patterns in massive open online courses for dropout reasoning and prediction," in *2016 IEEE Conference on Visual Analytics Science and Technology, VAST 2016, Baltimore, MD, USA, October 23-28, 2016*, 2016, pp. 111–120. [Online]. Available: http://dx.doi.org/10.1109/VAST.2016.7883517

[29] F. G. de Borja and C. M. D. S. Freitas, "Civisanalysis: Interactive visualization for exploring roll call data and representatives' voting behaviour." in *SIBGRAPI*. IEEE Computer Society, 2015, pp. 257–264. [Online]. Available: http://dblp.uni-trier.de/db/conf/sibgrapi/sibgrapi2015.html#BorjaF15

[30] K. Vaishnavi, A. Kannan, D. Cline, and R. Etemadpour, “A visualization tool for learning statistical analysis in multi tabular datasets,” *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, vol. 00, pp. 222–226, 2016.

[31] S. P. Cano, G. I. Alvarez, C. A. Collazos, and J. M. Arteaga, “A visualization and human-computer interaction proposal in the context of pronunciation information,” *IEEE-RITA*, vol. 11, no. 1, pp. 12–17, 2016. [Online]. Available: http://dx.doi.org/10.1109/RITA.2016.2518440

[32] H. Chernoff, “The use of faces to represent points in k-dimensional space graphically,” *Journal of the American Statistical Association*, no. 68, pp. 361–368, 1973.

[33] W. Zeng, C.-W. Fu, S. Müller Arisona, S. Schubiger, R. Burkhard, and K.-L. Ma, “Visualizing the Relationship Between Human Mobility and Points of Interest,” *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[34] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu, “Visualizing interchange patterns in massive movement data,” *Computer Graphics Forum*, vol. 32, no. 3pt3, pp. 271–280, 2013. [Online]. Available: http://dx.doi.org/10.1111/cgf.12114

[35] C. Musialek, C. Chen, H. He, S. Zha, and R. Johnson, “Traffic trender: Exploring bottlenecks at a systematic level.”

[36] M. M. Price, J. J. Crumley-Branyon, R. W. Leidheiser, and R. Pak, “Effects of information visualization on older adults’ decision-making performance in a medicare plan selection task: A comparative usability study,” *JMIR Hum Factors*, vol. 3, no. 1, p. e16, Jun 2016. [Online]. Available: http://humanfactors.jmir.org/2016/1/e16/

[37] M. Gusev, S. Ristov, and D. Mijoski, “Enhancing the educache simulator with visualization of cache performance,” in *2016 IEEE Global Engineering Education Conference, EDUCON 2016, Abu Dhabi, United Arab Emirates, April 10-13, 2016*, 2016, pp. 375–382. [Online]. Available: http://dx.doi.org/10.1109/EDUCON.2016.7474581

[38] M. Mamani, A. Infante, and A. Bergel, “Inti: Tracking performance issue using a compact and effective visualization,” in *33rd International Conference of the Chilean Computer Science Society, SCCC 2014, Talca, Maule, Chile, November 8-14, 2014*, 2014, pp. 132–134. [Online]. Available: http://dx.doi.org/10.1109/SCCC.2014.28

[39] M. Weber, R. Geisler, T. Hilbrich, M. Lieber, R. Brendel, R. Tschüter, H. Brunst, and W. E. Nagel, “Detection and visualization of performance variations to guide identification of application bottlenecks,” in *45th International Conference on Parallel Processing Workshops, ICPP Workshops*

2016, Philadelphia, PA, USA, August 16-19, 2016, 2016, pp. 289–298. [Online]. Available: http://dx.doi.org/10.1109/ICPPW.2016.50

[40] J. Nielsen, *Usability Engineering*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[41] J. Nielsen, "Thinking aloud: The #1 usability tool," January 2012, uRL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ [Retrieved March, 2018]. [Online]. Available: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

[42] C. Ware, *Information Visualization: Perception for Design*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004.

[43] J. Sauro, "Measuring usability with the system usability scale (sus)," February 2011, uRL: https://measuringu.com/sus/ [Retrieved March, 2018]. [Online]. Available: https://measuringu.com/sus/

[44] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*. John Wiley and Sons, 2003.