Exploring new ways of Usability testing for an E-Science/ Scientific research application

By

Julio Cesar Dovalina
St. Edward's University
Advisor – Dr. Gregory Madey

Abstract

There is a growing interest on the developing of better approaches to conduct usability testing on scientific research applications. This paper describes modified and different approaches and techniques of conducting usability testing to this relatively new infrastructure of E-Science/Scientific research applications. The infrastructure we will be focusing on this paper is intended to facilitate research and allow scientists to use computational resources in different physical sites and allow the remote access of the information source and simulation trough this Web based scientific research application. The NOM (Natural Organic Matter) simulation model which will be focusing on employs a diverse mixture of hardware, software, and information technologies such as J2EE (Java 2 Platform, Enterprise Edition), Microsoft's .Net, and web-based relational database management systems (RDBMS), which have made possible the use of the Web and network as a platform for this scientific research application.

We also discuss and analyze the different aspects of a scientific research application that should be considered when conducting usability testing on this type of applications. Because of the combination and interaction of aspects such as diverse range of technologies, audience expertise, functions feedback response, and tools and goal composition on a scientific research application, the approaches to be taken towards a research application are dissimilar from the type of usability testing techniques used for a commercial or an academic application. In addition to describing the different approaches and aspects that should be considered on a research application, we will also explore the differences between a scientific research application and a commercial piece of software and why is important to consider such differences when conducting usability testing on a research application.

1 Introduction

In the early start of the Web nobody really paid much attention to usability. In spite of everything, Web sites were more like collections of newspaper or magazine pages than they were similar to "real" software. The result of this line of thought was a lot of frustration. Getting the information you wanted right away was a random chance affair, with most sites doing a poor job of putting information in a form that users could understand [1].

In the 1990s, the interface development community started to develop and employ usability engineering methods to design and test software systems for ease of use, ease of learning, memorability, lack or errors, and satisfaction [1] [2] [3]. The World Wide Web was becoming a strong resource of information sharing and a popular medium for business advertising and transactions. It was no longer enough for the designers just to be sensitive to usability concerns. They had to set objective, measurable, operational usability goals [1]. Web usability gained strength and focus in the Human Computer Interaction community during the late 1990's [1] [4]. As the evolution of technology was allowing developers to develop new media style Web sites and applications, the number of usability problems was becoming even more complex, with a correspondingly greater negative impact on businesses revenues and customer retention [1] [5].

Now in the twentieth century, usability testing has become if not the most important one of the most important factors in developing a Web site and an application. In order for the users to use a Web site or a piece of software effectively, they must accomplish their tasks in the best way possible [6]. Studies of user behavior on the Web show a low tolerance for difficult designs or slow sites. People do not want to wait. People should be able to grasp the functionality of the Web site or software in the homepage or the main point of entry to the application [7]. Because of usability's impact between the success or failure of a system and how the lack of it can cost time, effort, and lead users to not buy or use a usable product, companies and usability practitioners in the last few years have started to employ new and more sophisticated techniques and approaches to improve their Web site's and application's usability.

2 Previous Work

During the late 1990s and the recent years, a variety of methods and approaches of usability testing have been developed. Some of the most commonly used include Ease of learning, Ease of use, User-centered design, Cognitive walkthrough, Focus groups, and Task analysis.

2.1 Ease of learning and Ease of use

In today's rapid changing technology usability testing is a necessity. How easy to perform a task or understand a Web site or system will dictate if a Web site or system fails or is successful. In addition, usability testing principles can be applied to computers, Web sites, and other software and for those reasons many new and old methods and processes have been renovated and developed for usability testing and measuring usability. In the late 1990s two main usability measures were used, Ease of learning and Ease of use. *Ease of learning* measures usability by comparing the time is takes a user to learn to do a job when working with an unfamiliar computer system or software to the time it takes them to learn to do the same job some other way [1]. *Ease of use* focuses on how many actions are required to complete a task successfully. An example will be the number of mouse clicks entered per procedure; the ease of use can be compared on this aspect on two different designs [1]. Both of these methods are still used today and even though they have changed since the late 1990s the main purpose remains the same, and they are still used today by big corporations to develop their sites and products [IBM]

2.2 User Centered Design

In this approach defining the user culture, including the user characteristics, user types, levels of expertise, and user task descriptions, are a prerequisite for interface development and testing [1]. The goal of this approach is to provide for the needs of all the potential users and target audience, adapting the technology and Web site to their expectations and avoid requiring them to go through unnecessary steps to complete a task and reach their goal [19]. This approach is mainly based on the three principles of Early human factor input, Empirical measurement of product usage, and Iterative design [8] [9].

The *Early human factor input* principle states that consideration should be given to human factors and user interface design guidelines very early in the process. *Empirical measurement of product usage* emphasizes on the behavioral measurements of ease of use and ease of learning previously mentioned [8] [9]. The *Iterative design* principle states that no matter how well in experience and background the first design is, designers should be prepared for a complete redesign, retest, and redesign otherwise the influence of iterative design becomes minimal [8] [9]. The user centered design approach is flexible and it can be applied to most Web sites and applications based on their structure its implementation may slightly change.

2.3 Cognitive Walkthrough

This approach evaluates how well the application supports exploratory learning and how difficult is for first time or infrequent users to operate the interface elements towards their goal [10]. The Cognitive walkthrough procedure requires some perquisites before implementing it, they include: general descriptions of the target users and the relevant knowledge they posses, a specific description of one or more representative tasks to be performed with the system, and a list of the correct actions to complete each of these tasks with the interface being evaluated. [11]. Cognitive walkthroughs can be performed at any stage of usability testing either to a design using a prototype, a conceptual design document, or a final product. This approach is basically used to help understand the usability of an application by users in an exploratory learning mode [10].

2.4 Task Environment Analysis

The task analysis approach is used to determine functionality by distinguishing the tasks and subtasks performed by the user [1]. Particular focus is paid to the user's frequent tasks, occasional tasks, exceptional tasks, and errors. According to Badre, part of a good task analysis is to identify strategies (combinations of tasks) used to reach the user's goals. With the rapid increase of new technologies, and the power of representation and simulation technologies, Web design will require us to cover all aspects of the Web task environment, including the physical, social, and visual [1]. The basic purpose of this method is to evaluate how people actually accomplish their goals with a piece of software [1].

2.5 Focus Groups

Focus groups are mainly used for exploring and discovering information, and they are typically used before development [12]. This method is not part of usability testing but it is a reasonable way to define usability metrics, or measures, to be used later in usability testing [12]. On interactive systems development, the proper role of focus groups can be used to discover what users want from the system. Focus groups are at some extent an informal technique that can help you assess user needs both before interface design and long after implementation. Even though focus groups may be a good way for exploring and discovering user needs, focus groups are a rather poor method for evaluating interface usability, thus is not a good idea to rely just on them as your only method in a Web design or application developing. [13].

The methods and principles mentioned above are commonly used today as tools to conduct usability testing. In combination, all of these different methods can serve as excellent tools for conducting usability testing, measuring it, and evaluating it.

3 Exploring usability testing involving E-Science and a scientific research application

For large scale web based scientific research applications usability testing is very important since the purpose of such application is to facilitate research ,allow collaboration among its users, share personal experience, and accelerate the dissemination of knowledge [14]. If a research application is not usable and reliable it will not fulfill its purpose. Web based research applications use a diverse range of software, hardware, and information technologies and this can become an obstacle for the users when using the application [15]. In addition, the differences and contrasts between a scientific research application and a typical Web site or E-commerce site can become a problem when using the typical usability testing methods. For that reason such differences need to be explored and identified towards the goal of developing a usability test plan that tests and evaluates all the aspects that are important to a research scientific application.

Exploring and identifying the differences between a research application and a commercial one is crucial. The current usability testing methods and techniques lean towards testing for consumer satisfaction and making their application and Web sites more usable for consumers in order for them to feel more conformable and willingly to buy their products. If we apply those same techniques towards a research application the results will more likely show a lack of focus on aspects such as device diversity, audience expertise, function's feedback response, tools and goal composition, aspects that are very important to be taken into consideration when conducting usability testing on a scientific research application.

3.1 Purpose

A scientific research application main purpose is to facilitate research, allow collaboration among its users, share personal experience, and accelerate the dissemination of knowledge [14]. On a research application the user should not expect much in return without much work and effort. For example, the NOM simulator it's a Web based scientific research application that facilitates research by simulating the complex behavior of the Natural Organic Matter molecules over time [14]. Research applications facilitate research and collaboration among its users, but in order for the users to get significant feedback they need to put a substantial amount of effort and work in return. A scientific research application facilitates research, but such application will not serve its purpose if the user does not put a considerable amount of effort and work in return.

On the other hand, on an E-commerce application the features and components are designed to mainly think for the user to focus the user's attention on the services and products that they offer. On most well designed E-commerce applications the users do not have to put much effort or work to get feedback, a service, or a product.

Therefore, if we use typical usability approaches without exploring the aspects that are important for a research application, the influence of any approaches used will be minimal. The different aspects regarding a research application play an important role when conducting usability testing and they should be taken into consideration to accomplish a well structured usability testing plan.

3.2 Device Diversity

A web based scientific research application such as the NOM simulator is composed of a diversity of hardware and software. Such diversity of technologies is needed in most scientific research applications to be a good research facilitator. In the case of the NOM simulator, a diversity of technologies were required to facilitate research and allow the users to collaborate and do research trough a Web site from different physical sites and have remote access to the simulation information.

With Web systems and their mixture of hardware to support, the environment can become very difficult to control [15]. For example, the NOM simulator uses a mixture of hardware that involves a Java 2 Enterprise Edition architecture running on a Linux 8.0 and Windows 2000 operating systems. On the software part, the machines on the cluster include a HTTP server, application servers, database servers, a reports server, and a data mining server. The architecture is showed in Figure 1.1 [14].

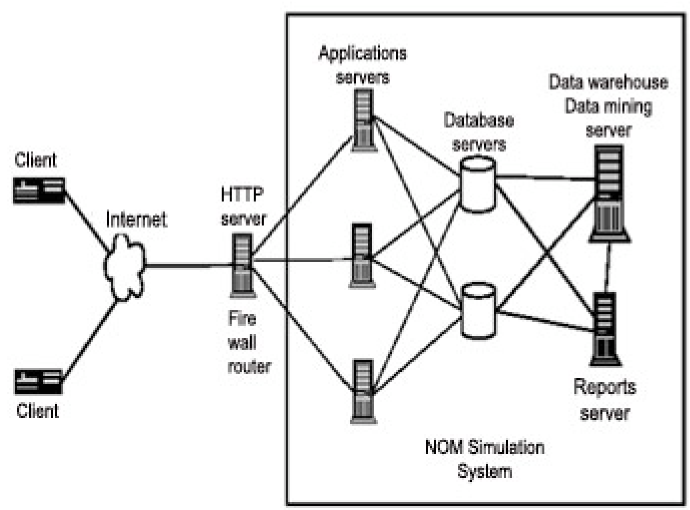

Figure 1.1 The architecture of NOM simulation Model

The Web interface is written in Java Server Pages, Java Servlets, and Java Beans that communicate with a database via Java Database Connection (JDBC). For the design of the Web interface, standard HTML, XML/XSL style sheets, and JSPs were used. The implementation model for the Web interface is shown in figure 2.1 [14].

The NOM simulator is also composed of different features and intelligent components which also use mixed technologies on their own. This wide range of different technologies can become difficult to control and use. In addition, the Web has serious constrains in terms of connection speeds and browser display, both which can impact directly impact the Web experiences from user to user [15]. Thus, the usability testing plan should include quality assurance across devices, operating systems, browsers, and modem/LAN connections [16]. Especially since the purpose of a research application is to facilitate research, the optimal performance and reliability of the application is needed in order for the users to fully take advantage of its capabilities.

An aspect such as device diversity cannot be omitted when conducting usability testing, with such diversity of technologies in use the omission of this aspect might lead a an unstable and not very practical and usable application.

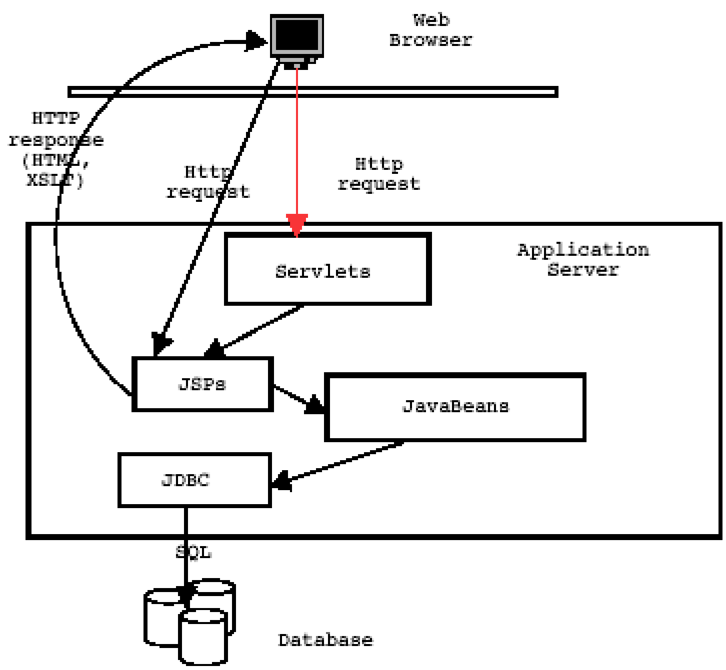

Figure 2.1 The implementation model for the Web interface

3.3 Audience Expertise

In most scientific research applications the user is expected to have a considerable level of expertise on the research topic. A research application is not similar to a commercial one where the user is not expected to know much about the content or purpose of the page to utilize it. On the other hand, when using a research application the user needs a reasonable level of expertise on the research topic otherwise it may be very difficult or almost impossible for the user to obtain significant feedback and take advantage of the applications research tools.

For example, to be able to efficiently use the NOM simulator the user needs a reasonable level of understanding of Natural Organic Matter. Since the research being conducted involves the complex behavior of the NOM over time, the lack of understanding of this molecule will not allow the user to go further on the application or get any valuable feedback. The user will encounter many problems understanding the purpose of the application and it's potential.

This difference on levels of expertise expectations of the users between a commercial and a research application makes the difference when conducting usability testing. For example, if it is decided that the user centered design will be used, the prerequisite used by this approach regarding the level of expertise should be meticulously taken into consideration. If this aspect of a research application is not researched well enough while conducting usability testing, the testing influences will be minimal and the application's users and the application itself will fail to meet its goal.

On a cognitive walkthrough approach the same rule applies. Given that on this approach the user's relevant knowledge is also considered, when using this approach more emphasis should be put on this aspect because if omitted it may lead to an application with poor usability for any type of audience in general. Most research applications have a target audience that will require certain level of expertise from the users to be able to use and take advantage of the application. But if the tester fails to consider this aspect while conducting usability testing, the application may not become a very practical and usable tool.

As it can be seen for the reasons mentioned above, the omission of the audience expertise aspect may cause a research application to not meet its goals and fail as a research facilitator tool. This aspect on its own in many cases may dictate the applications failure or success.

3.4 Function Feedback Response

Scientific research applications such as the NOM simulator are composed of a variety features and intelligent components that are designed to facilitate research, and the feature's feedback response might not be as fast as commercial applications. In the case of the NOM simulation model, the main purpose of this research application is to simulate the behavior of the NOM molecules over time [14]. In order for the user to obtain valuable feedback from the application they have to wait for long periods of time, it can take hours, days, weeks, and even months if they want to view significant changes on the molecule's behavior over time.

How quick the application's features give you feedback is not major issue when conducting usability testing in most research applications compared to a commercial piece of software where it is almost required that you get an immediate response or feedback when using one of its features to obtain a service or product. Research application's tools are created and designed to help facilitate research where in most cases it can take a considerable amount of time to obtain valuable feedback and results.

The aspect of how fast a function responds to a user's input should be considered, however, it should not be a major consideration when conducting usability testing on a research application since research and commercial applications greatly differ in purpose.

Instead, the test should consider how the application tools facilitate research and provide quality and reliable feedback and results.

3.5 Tools and Goal Composition

This section will describe a technique proposed by Jakob Nielsen intended to extend task analysis based on the goal composition principle suggested by Clayton Lewis. In simple terms, goal composition begins by considering each primary goal that the user may have when using an application [17] [18]. Then a list of possible additional features is then generated by combining each of these goals with a set of general meta-goals that extends the primary goals [17] [18].

We focused on only two of the main categories of goal composition, *generalization mechanism* and *user control mechanism* and five of the principles within them which are *multiplexing and reuse* within *generalization mechanism and monitoring, recording and retrieving, modification and editing* within *user control mechanism*.

After observing the NOM simulation model and the interaction between its features and intelligent components; I have concluded that the principles mentioned above in the developing and testing stages of a scientific research application if applicable can become a useful tool to conduct usability testing. If applicable, the developer and usability testers may be able to use these principles when considering the future of the application's features potential to evolve and the creation of new ones when combining the current feature's goals and subgoals as proposed by this approach.

The purpose of a research application is to facilitate research, but it is also intended to evolve according to the user's goals and needs and by applying some of these principles the process may become more feasible to conduct.

3.5.1 Generalization Mechanism

The Generalization Mechanism is intended to allow users to increase the scope of each feature by applying it to more objects [17] [18].

Multiplexing refers to the ability to have many instances of a goal be achieved as a single goal [Lewis, Nielsen]. For example, when using the NOM simulator to configure a new simulation with particular parameters, the user might want to know if a simulation with similar parameters to the ones he/she is configuring has already been created. In the case of the NOM simulator such functionality while configuring a new simulation has already been created to let the user know before invoking a new simulation if there is a configuration similar to the one he/she is trying to invoke. Multiplexing support was shown by the functionality of two features; *create a new simulation* and *finding a similar simulation* combined and working together to achieve the goal of configuring a new simulation but achieving instances of a goal such as finding a similar one in the process.

Reuse enables the user to utilize part of the work towards one goal when wanting to achieve a similar goal [17] [18]. For example, the NOM simulator users may want to be able to reuse the information of previous simulation configurations and be able to modify them to invoke a new one. Those functionalities and features were considered by the developers while developing this particular research application. The users have the capability of uploading and downloading XML files that defines one or more molecule types and define a set of environment parameters where a molecule id and an environment id are returned to the user for further reference. The user can also upload or download an XML file, modify the configuration parameters, and invoke a simulation [14]. In addition, the users are capable of using reports of previous configurations to decide what parameters to use in a new one. Reuse may accordingly often take place in combination with the recording a retrieving mechanism which is discussed below [17] [18].

In order to have an efficient and facilitating research application, it is recommended that user has the option to reuse previous data and procedures of previous goals towards new goals to avoid having to recreate a whole new procedure when having similar data that can be reused and modified. Also the multiplexing of features is a useful principle that may be considered when thinking on avoiding the user having to go trough unnecessary steps and tasks to accomplish several goals that can be accomplished by a single procedure.

3.5.2 User Control Mechanism

This mechanism emphasizes on allowing the user to inspect and change the way the computer carries out their instructions [17] [18].

The *monitoring* refers to the user's capability of determining whether their goal has been achieved or what progress is being made towards their goal when asking an application to perform a task [17] [18]. It is important that a research application provides the user with options to preview the progress towards their goal and be notified when a determined task has been achieved successfully. Having such options will facilitate the user's research by allowing them to perform other tasks simultaneously and be able to keep track of each of them. For example, in the case of the NOM project the users have the ability of viewing running time reports, which provides the user with information of the current behavior of the molecules at the time. In addition, the user receives an email each time they invoke a new simulation to let them know that the simulation is executing and one to let them know the status of the simulation, either it was successfully completed or it was terminated due to any exception by hardware malfunction or software problem [14]. This is possible with the help of a sending email and report intelligent components [14].

Recording and Retrieving refers to the user's need to keep record of what has been accomplished [17] [18]. In the case of a scientific research application it is fundamental that the user be capable of keeping track of the data and results being produced by the application and their inputs. In a research application recording and retrieving previous and current data is a primordial element since the user should be able to compare and analyze the data to allow the possibility of discovering new patterns and be able to better understand it. For example, in the NOM project the users are provided with tools such as simulation reports and XML uploading and downloading of files that allow the user to keep track of configuration information. [14]

The *Modification and Editing* principle refers to the ability of the user to modify the outcomes of prior operations in addition to canceling or starting over an operation.

It is an important aspect for a scientific research application to provide the user with options to cancel, start over, modify, and edit the outcomes of previous and current operations. This will help the researcher avoid unnecessary tasks to recreate a previous operation. For example, in the NOM project the user has the option of terminating a simulation once it has been invoked, that option avoids extra stress on the system and database by terminating unnecessary simulations. The user also has the option of decreasing and increasing the simulation time once a simulation has been invoked [14].

Modification and editing, recording and retrieving, and monitoring are principles that are ought to be considered when developing and testing the usability of a research application. These principles touch important aspects such as being able to determine the progress of the user action's, keeping track of the data being produced, and allowing the user to cancel, start over, modify, and edit the outcomes of previous and current operations. All of these are vital aspects that should be considered when conducting usability testing if the goal is to facilitate research by creating a usable and practical research application.

4 Conclusion

In this paper several approaches and techniques for conducting usability testing on a scientific research application have been presented. We also presented some of the major differences between a commercial application and scientific research application and why such differences must be considered when conducting usability testing on a research application.

The approaches and techniques mentioned on this paper focused on the NOM simulation model, but this model possesses the characteristic which typical scientific applications have. Therefore, the usability testing techniques recommended on this paper can generally be applied to other scientific research applications.

There are many possible extensions to the approaches and techniques covered on this paper. Many of the approaches analyzed on this paper are composed of principles that were not mentioned, which may be applicable to other research application.

Bibliography

[1] Badre N. Albert. *Shaping Web Usability – Interaction Design in Context.* Addison-Wesley, pages 5, 6, 7 2002.

[2] Gould, J. D., and Lewis, C.H. *Designing for usability. Key principles and what designers think.* Communication of the ACM, pages 28,300-311, 1985

[3] Nielsen, Jakob. *Usability Engineering*. New York, Academic Press 1993.

[4] Nielsen, Jakob. *Alertbox, why you only need to test with 5 users*. New York, Academic http://www.useit.com/alertbox.980809.html, 2000.

[5] Manning, H., McCarthy, J.C., and Souza, R.K. *Why most Web sites fail.* Technology Series, 3(7), 1998

[6] What is Usability? http://www.usabilityfirst.com/intro/index.txl, 2002

[7] Nielsen, Jakob. *Why Is Usability Important?* http://usability.gov/basics/index.html#importance, 2002.

[8] Rubin, Jeffrey. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wisely Technical Communication Library, pages 10-13, 2003.

[9] Gould, J.D. and Lewis, C.H.. *Designing for usability: Key Principles and what designers think*. Communication of the ACM, pages 28, 300-311, 1985.

[10] *Methods:Cognitive Walkthrough.* http://usabilityfirst.com/methods/cogwalk.txl, 2002.

[11] John Rieman, Marita Franzke, and David Redmiles. *Usability Evaluation with the Cognitive Walkthrough*. http://turing.acm.org/sigs/sigchi /chi95/proceedings/tutors/jr_bdy.htm, 1995.

[12] Barnum, Carol M. *Usability Testing and Research.* Longman, pages 44-45, 2002.

[13] Nielsen, Jakob. *The Use and Misuse of Focus Groups* .http://useit.com –papers and essays – focus groups. 1997

[14] Xiarong Xian. *Agent-Based Scientific Applications and Collaboration Using Java*. Master's thesis, University of Notre Dame, 2002.

[15] Nguyern Q. Hung. *Testing Applications on the Web- Test Planning for Internet Based Systems.* Wiley Computer Publishing, pages22, 23, 2001

[16] Nielsen, Jakob. *The Difference between Web design and GUI design.* http://www.useit.com/alertbox/9705a.html, 1997.

[17] Jakob Nielsen. *Goal Composition: Extending Task Analysis to Predict Things People May Want to Do*. http://useit.mondosearch.com/cgi bin/MsmFind.exe?QUERY=Goal+Composition, 1994.

[18] C.H. Lewis. *A research agenda for the Nineties in human computer interaction*. Human Computer Interaction, 5(2-3):125--143, 1990. 25

[19] *User-centered design*. http://webstyleguide.com/interface/user-centered.html. 2003