Usability and Reliability of the User Action Framework: A Theoretical Foundation for Usability Engineering Activities

Sriram Sridharan

Thesis submitted to the Faculty of Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Industrial and Systems Engineering

Tonya L. Smith-Jackson, Ph.D., Co-Chair
H. Rex Hartson, Ph.D., Co-Chair
Robert C. Williges, Ph.D

Keywords: User Action Framework, Usability Engineering, Web Usability,
Formative Usability Evaluation

Usability and Reliability of the User Action Framework:
A Theoretical Foundation for Usability Engineering Activities

Sriram Sridharan

ABSTRACT

Various methods exist for performing usability evaluations, but there is no systematic framework for guiding and structuring assessment and reporting activities (Andre et al., 2000). Researchers at Virginia Tech have developed a theoretical foundation called the User Action Framework (UAF), which is an adaptation and extension of Norman's action model (1986). The main objective of developing the User Action Framework was to provide usability practitioners with a reliable and structured tool set for usability engineering support activities like classifying and reporting usability problems. In practice, the tool set has a web-based interface, with the User Action Framework serving as an underlying foundation.

To be an effective classification and reporting tool, the UAF should be usable and reliable. This work addressed two important research activities to help determine the usability and reliability of the User Action Framework. First, we conducted a formative evaluation of the UAF Explorer, a component of the UAF, and its content. This led a re-design effort to fix these problems and to provide an interface that resulted in a more efficient and satisfying user experience. Another purpose of this research was to conduct a reliability study to determine if the User Action Framework showed significantly better than chance agreement when usability practitioners classified a given set of usability problem descriptions according to the structure of the UAF. The User Action Framework showed higher agreement scores compared to previous work using the tool.

DEDICATION

I dedicate this work to my father, S. Sridharan; he never ceases to inspire me with his hard work and desire to learn more.

ACKNOWLEDGEMENTS

This thesis has afforded me an opportunity to thank many individuals that helped me during the course of my graduate student life. First and foremost, I would like to thank my parents, who were very encouraging when I wanted to come to the US for my Masters education. Being an international student poses many challenges, any my parents always provided me moral and financial support, to overcome those difficulties.

I am very grateful that I was blessed with a wonderful committee. My co-chairs, Dr. Tonya L. Smith Jackson and Dr. H. Rex Hartson, offered invaluable guidance during the entire route of this research. In spite of her busy schedule, Dr. Smith Jackson took time to meet with me whenever I knocked on her door. She always showed me the correct path when it comes to conducting research and scientific writing. I cannot thank her enough for being by my side whenever I needed support and direction. As a co-chair from the computer science department, Dr. Hartson provided the initial idea and motivation to pursue this research. We had numerous discussions and brainstorming sessions, which helped shape this work and provided me a clear idea of what was expected out of my thesis. I am very thankful that I will be immensely benefited by his approach to solve HCI problems, in the years to come. Many thanks to Dr. Robert C. Williges who, with his abundant experience, helped me conduct the research in an organized manner and cleared any doubts fogging my mind during the course of this work.

Thanks to Dr. Robert S. Schulman for explaining kappa calculations and Dr. Terrence Andre for sharing his data sheets from his work. Thanks to my colleague, Miranda Capra, who helped me set up the usability laboratory and gave her opinion whenever I needed any clarification. I am also thankful to Priya Shivakumar, who helped create my own copy of the UAF and, for pulling those re-design sessions. I thank all my participants that helped me with this research and endured those long laboratory sessions. Lastly, many thanks to my friend Bama Ranganathan, who patiently proof read my documents and, for being there to share my moments of joy and sorrow.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

PROBLEM STATEMENT

Researchers in Virginia Tech are building a suite of usability engineering support tools based on a theoretical framework of concepts and issues, called the User Action Framework - UAF (Keenan, 1996; Van Rens, 1997; Andre, 1999). The UAF is a theory based, interaction-style-independent structured knowledge base of usability issues and concepts. The components of this suite of tools consist of UAF Explorer, Usability Problem Classifier, Usability Problem Inspector and Database Manager. Of these the UAF Explorer is in working stage while the other components of the tool set are in the development stage (H. R. Hartson, April-2001, Personal Communication). In practice, these components of the tool set have a web-based interface, with the UAF serving as an underlying foundation. Lee (2000) observes that in order to make effective use of the Web, web site designers need not only provide useful content, but also present that content in a way that results in satisfying user experience for specific tasks. This user experience has become the definition of "usability" (ISO, 1993). The intention is to test the usability of the web-interface of the UAF Explorer and its content by means of a formative evaluation study. Formative evaluation is defined as the evaluation of interaction design as it is being developed, early and continually throughout the interface development process (Carroll, Singley & Rosson, 1992). The fact that the existing web interface has not been tested for its usability stresses the need for such an evaluation at this stage.

One important performance measure of a usability-engineering tool is reliability, which is a measure of consistency, or the extent of agreement, among evaluators with respect to their results in using the tool (Andre et al., 2000). High reliability means the usability practitioners share a clear understanding of the structure and content of the user action framework and this is essential for its role as a common foundation for the suite of UAF tools. Without this kind of reliability, the usability data will depend only on the individuals using the tool. Consistent understanding and reporting of underlying causes of usability problems by usability experts is important for cost-effective analysis and re-design. Otherwise, one of the very intents of the UAF, namely facilitating high-quality and consistent usability problem reporting will not be achieved. Also, the latest version of the User Action Framework features some changes in terms of its quasi-hierarchical structure and content since the reliability study done by Andre (1999). Due to these changes it was necessary to study the reliability to see if the UAF still has good reliability.

RESEARCH GOALS

The goals of this research from the problem statement are as follows:

1. To conduct a formative usability evaluation of the web-based interface of the UAF Explorer, a component of the User Action Framework, and its content. The purpose is to uncover as many usability problems this interface might have and to fix them so that the usability will not be a confounding factor in the reliability study.
2. To conduct a reliability study to determine whether the User Action Framework shows significantly better than chance agreement when usability practitioners classify a given set of usability problem descriptions using the UAF.

The work accomplished through these research goals was used to develop re-design recommendations based on the formative usability evaluation study. These design recommendations led a re-design effort that fixed the usability problems found during the study. Also, this research demonstrated the reliability of User Action Framework facilitating usability practitioners to use the User Action Framework as a reliable foundation for usability engineering activities.

BACKGROUND - USER ACTION FRAMEWORK

As mentioned earlier, there is a lack of systematic framework for guiding and structuring assessment and reporting activities, although various methods exist for performing usability evaluation (Andre et al., 2000). To solve this problem, researchers in Virginia Tech have developed a suite of usability engineering support tools based on a theoretical framework of concepts and issues, called the User Action Framework (UAF). As an adaptation and extension of Norman's action model (1986), the UAF is based on what the users think, perceive and do throughout each cycle of interaction with a machine. Each category on the UAF reflects Norman's stages of action model and has a usability issue associated with it (e.g.: noticeability, clarity etc). The UAF has a quasi-hierarchical structure, with mutually exclusive sub-categories under each category.

User Action Framework – Terminology

Each category (part of the interaction cycle) of the User Action Framework will have one or more of the descriptive information as given below.

Node Content (Category)

The node content for explains the meaning and purpose of that node each component of the interaction cycle.

Sub-Categories

The UAF structure is quasi-hierarchical. i.e., it is mostly hierarchical, with mutually exclusive sub-categories under each category. However, occasionally more than one path leads to a given node, where the order of choices is immaterial.

Cross Reference

Cross-references are descriptions of cases that might seem to belong in this category, but in fact do not. Also, cross-references gives reasons why a particular case does not belong there, and some help about where one should look for such cases.

User Action Framework – Content

Each category of the User Action Framework has been explained in terms of what it means in interaction activity. An overview of each of the interaction activities has also been given in tabular form.

Planning

Planning occurs when users determine what actions need to be taken and how to go about completing those actions. It is concerned with the user's ability to understand the overall computer application in the perspective of work context, problem domain, environmental requirements and constraints. The primary focus is on the system model and metaphors, and the users' knowledge of system state and modalities. Planning includes user work goal decomposition across a hierarchy of plan entities: goals, tasks, and intentions.

Table 1.1: Overview of Planning

Category – Interaction activity	**Associated usability issues**	**Primary Sub-Categories**
Planning: User deciding what to do	• Metaphors, user's model of system, task planning and decomposition	User's model of system
		Goal decomposition
		Supporting planning for error avoidance
		User's knowledge of system state, modalities, and especially active modes
		User and work context
		User's ability to keep track of how much is done

Translation

Translation is about cognitive affordances to support the users' ability to plan physical actions. The user draws on knowledge, experience, and cognitive affordances in the interaction design to establish, or determine an action plan to carry out the intention. The translation issue accounts for a large proportion of usability problems observed in the field and is purely cognitive in that the user has formed a mental plan for actions, but has not yet done those actions. Usability issues in the UAF under translation include those that pertain to presentation of the cognitive affordances (e.g., perceptual issues, legibility, noticeability, timing etc). Translation issues also include effectiveness of content or meaning of cognitive affordance (e.g., issues of clarity, completeness, error avoidance, consistency etc)

Table 1-2: Overview of Translation

Category – Interaction activity	**Associated usability issues**	**Primary Sub-Categories**
Translation: User determining how to do it, translating plans into actions	• Existence of cognitive affordance (e.g., visual cue)	Existence (of a way or of a cognitive affordance to show the way)
	• Presentation of cognitive affordance (e.g., noticeability, legibility, layout and grouping)	Presentation (of a cognitive affordance)
	• Content/meaning of cognitive affordance (e.g., clarity, completeness, correctness, relevance)	Content, meaning (of a cognitive affordance)
	• Task structure and interaction control (e.g., locus of control, direct manipulation, cognitive directness)	Task structure and interaction control
	• Preferences and efficiency (e.g., number of steps, short cuts, anticipating most likely next tasks, preference settings etc)	Preferences and efficiency (of Translation issues)

Physical Action

Executing the planned actions is the focus of the physical action part of the interaction cycle, which is about perceiving objects to manipulate objects. Perception has to do with the usual factors of noticeability, legibility, contrast, and timing. Object manipulation has to do with interaction complexity, input/output devices, interaction styles and techniques, manual dexterity, layout (Fitts' law), and physical disabilities.

Table 1-3: Overview of Physical Action

<table>
<tr><th>Category – Interaction activity</th><th>Associated usability issues</th><th>Primary Sub-Categories</th></tr>
<tr><td rowspan="2">Physical actions: User doing the actions</td><td rowspan="2">Fitts' law factors
Awkwardness and fatigue
Physical disabilities
Power performance for experts</td><td>Perception of manipulable and manipulated objects</td></tr>
<tr><td>Manipulating objects</td></tr>
</table>

Assessment

The assessment portion supports the users' ability to gauge the outcome of physical actions. It parallels the translation part in that it has to do with presentation of feedback, meaning of feedback, and preferences and efficiency. The assessment portion contains the usual issues of legibility, noticeability, timing, layout, grouping, presentation of feedback and also involves complexity, clutter, consistency, organization of information displays, and presentation medium.

Table 1-4: Overview of Assessment

Category – Interaction activity	**Associated usability issues**	**Primary Sub-Categories**
Assessment: User assessing outcome	• Existence of feedback • Presentation of feedback (e.g., noticeability, legibility, timing, presentation medium) • Content/meaning of feedback (e.g., clarity, completeness, correctness, relevance)	Issues about feedback (about interaction for task) Issues about information displays (results for task)

Independent

The independent portion of the User Action Framework contains usability issues and overall characteristics not related to a specific part of users' interaction cycle. It includes the overall look and feel issues and also issues like consistency, use of graphics and sound etc.

Table 1-5: Overview of Independent

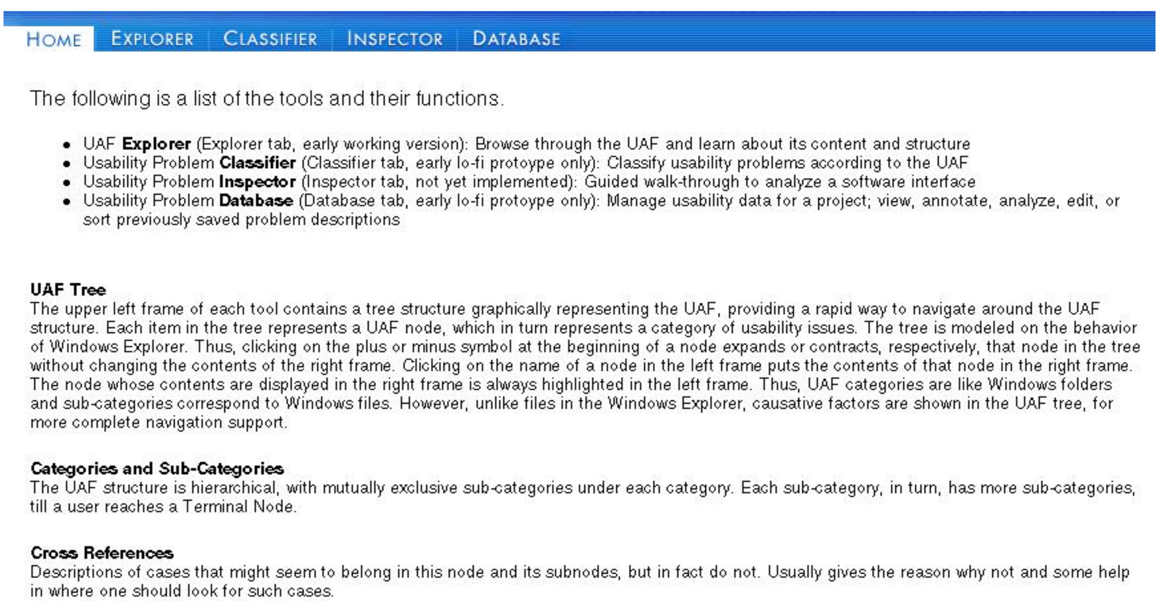
Category – Interaction activity	**Associated usability issues**	**Primary Sub-Categories**
Independent: of place in Interaction Cycle	• Look and feel - Aesthetics • Consistency • Overall Style	Overall Look and Feel Portability Overall style Overall interaction complexity Overall use of direct manipulation Overall preferences and efficiency issues Overall use of modes and modalities

User Action Framework – Interface and How it Works

In practice, the tool set has a web-based interface, with the User Action Framework serving as an underlying foundation. The interface of the UAF Explorer, the main focus of this formative usability evaluation, has been explained in detail and an effort has been made to briefly describe the other components of the tool set for the sake of completeness.

UAF Home Page

Figure 1-1 shows the home page of the web-based suite of tools. The purpose of this page is to give a brief introduction about the User Action Framework, and to explain what the User Action Framework does in terms of interaction design. The home page also gives a list of tools that are available and briefly explains how to use the tool set to explore the User Action Framework in order to learn about its structure / content and to navigate between various tools.


HOME EXPLORER CLASSIFIER INSPECTOR DATABASE

The following is a list of the tools and their functions.

- UAF **Explorer** (Explorer tab, early working version): Browse through the UAF and learn about its content and structure
- Usability Problem **Classifier** (Classifier tab, early lo-fi protoype only): Classify usability problems according to the UAF
- Usability Problem **Inspector** (Inspector tab, not yet implemented): Guided walk-through to analyze a software interface
- Usability Problem **Database** (Database tab, early lo-fi protoype only): Manage usability data for a project; view, annotate, analyze, edit, or sort previously saved problem descriptions

UAF Tree
The upper left frame of each tool contains a tree structure graphically representing the UAF, providing a rapid way to navigate around the UAF structure. Each item in the tree represents a UAF node, which in turn represents a category of usability issues. The tree is modeled on the behavior of Windows Explorer. Thus, clicking on the plus or minus symbol at the beginning of a node expands or contracts, respectively, that node in the tree without changing the contents of the right frame. Clicking on the name of a node in the left frame puts the contents of that node in the right frame. The node whose contents are displayed in the right frame is always highlighted in the left frame. Thus, UAF categories are like Windows folders and sub-categories correspond to Windows files. However, unlike files in the Windows Explorer, causative factors are shown in the UAF tree, for more complete navigation support.

Categories and Sub-Categories
The UAF structure is hierarchical, with mutually exclusive sub-categories under each category. Each sub-category, in turn, has more sub-categories, till a user reaches a Terminal Node.

Cross References
Descriptions of cases that might seem to belong in this node and its subnodes, but in fact do not. Usually gives the reason why not and some help in where one should look for such cases.

Figure 1-1: Home page of the User Action Framework

UAF Explorer

There are two components to the UAF Explorer (Fig 1-2) – the Explorer, which is the frame at the right-hand side and the Explorer Toolbar, which is seen as a frame in the lower left-hand side of the screen-shot. Usability practitioners will use the UAF Explorer to navigate and access the basic tools and the interaction style content of the UAF. Exploring the UAF using the Explorer is a way for a usability practitioner to learn about the issues and concepts and can be considered as an "entry-point" to use the entire tool set. Using the explorer, the users will be able to learn all about the UAF – about its quasi-hierarchical structure, about its content and about different terminologies that will be used in the tool set. Also, individuals can use the Explorer Toolbar to search the UAF and add annotations to their copy of the UAF to tailor it to their own special needs.

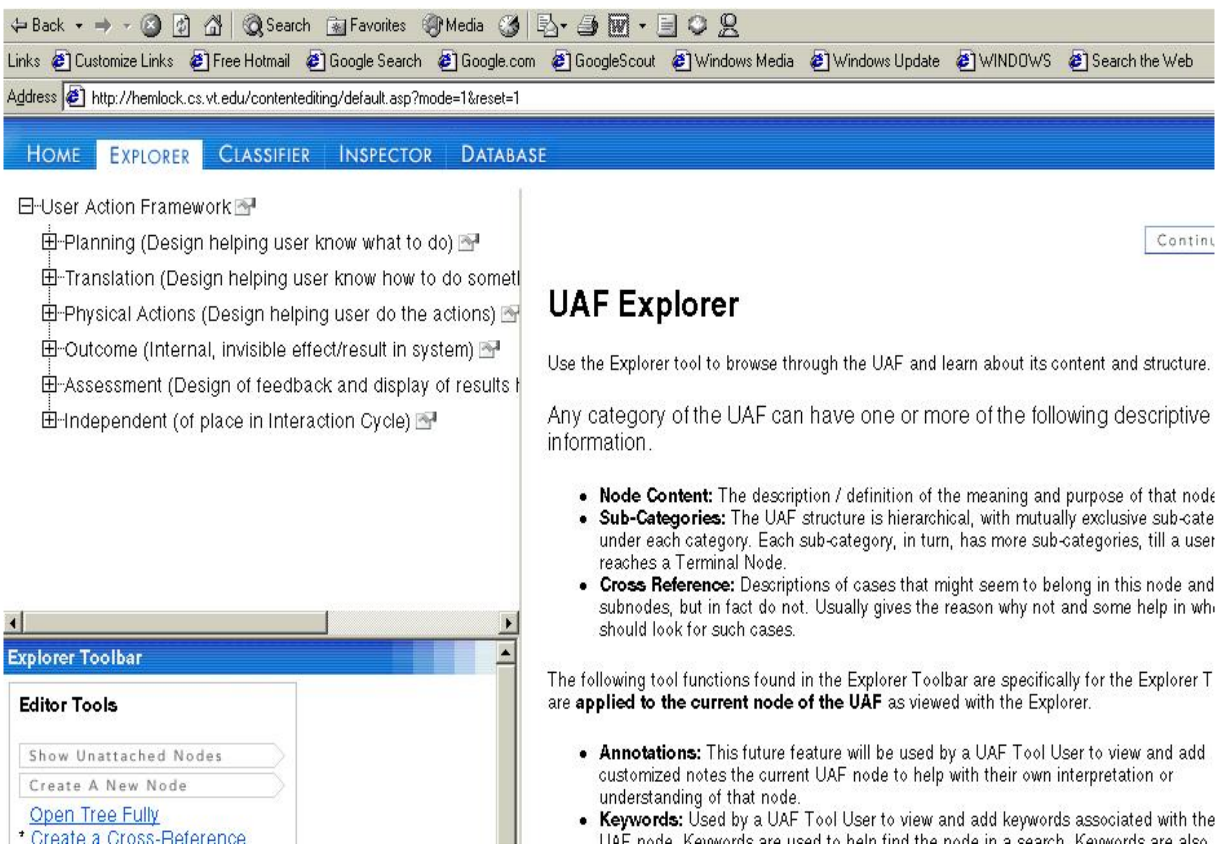


Figure 1-2: Screen shot of the UAF Explorer tool

The left frame of the Explorer tool changes the contents of the right frame. Clicking on the name of a node in the left contains a tree structure graphically representing the User Action Framework, providing a rapid way to navigate around the UAF structure. Each item in the tree represents a UAF node, which in turn represents a category of usability issues. The tree is modeled on the behavior of Windows Explorer. Thus, clicking on the plus or minus symbol at the beginning of a node expands or contracts that node in the tree, and places the contents of that node in the right frame. The node whose contents are displayed in the right frame is always highlighted in the left frame. Thus, User Action Framework categories are like Windows folders and the sub-categories correspond to Windows files. However, unlike files in the Windows Explorer, the sub-categories are shown in the User Action Framework tree, for more complete navigation support. By selecting one of the UAF nodes, the user will start traversing the structure as shown in Figure 1-3.

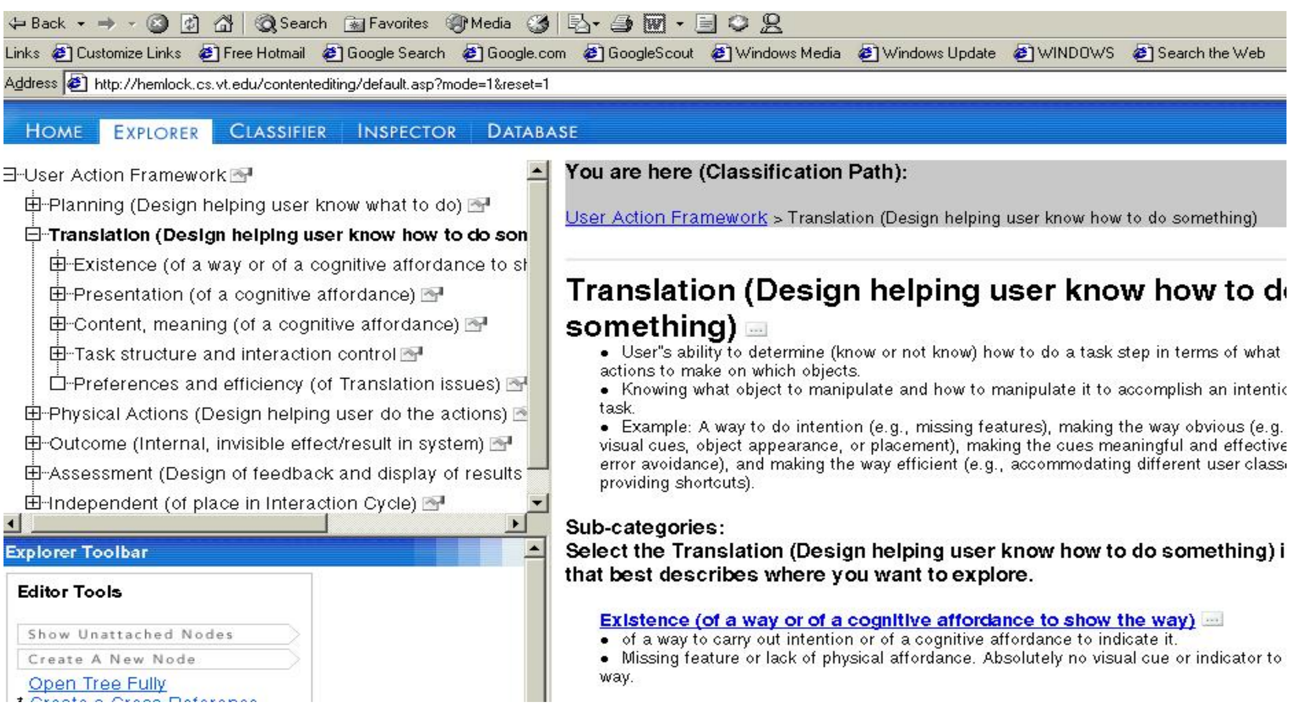


Figure 1-3: An example of how a user would traverse the UAF

Usability Problem Classifier

Practitioners will use the Usability Problem Classifier to classify, describe, and report observed usability problems according to the UAF content. Since the UAF is a quasi-hierarchical structure of usability attributes, practitioners traverse the UAF as a decision structure to classify an identified usability problem, selecting the most appropriate classification category and sub-category at each level of the hierarchy, and locating the problem in a structured space of usability problem types. The cumulative set of category choices along the classification path is taken as a sequence of usability attributes, a kind of "encoding" that determines a complete classification description of the usability problem in question.

Usability Problem Inspector

Usability inspection is a cost-effective alternative or supplement to lab-based formative usability evaluation at any stage of development (e.g., design sketches, prototypes, or fully implemented systems). The Usability Problem Inspector in the UAF is an inspection tool to be used by a usability specialist in the role of usability evaluator. A usability inspector using the Usability Problem Inspector is driven by inspection questions that come from each node of the UAF structure. If the entire UAF content is used, the Usability Problem Inspector will be capable of a very broad and thorough (but lengthy) usability inspection. However, typically only parts of this broad scope will be selected to match the specific needs of a given inspection instance more cost effectively. This selection is accomplished through the use of an inspection filter, essentially a database query, which acts as a declarative statement of question relevance. During inspection, the Usability Problem Inspector traverses the hierarchical structure of the UAF – first for planning, then for translation, physical actions, and assessment, starting at a high level and working down to details – visiting each node one-at-a-time. Only relevant inspection questions are shown to the practitioner according to whether each node visited matches the filter, tightly tailoring the scope to the specific needs of an inspection instance.

Database Manager

Practitioners will use the Usability Management tool, with its underlying relational database keyed to UAF categories, to create a project record of each usability problem accompanied by such information as severity rating, alternative solutions, estimated cost to fix, cost/importance analysis, priority-to-fix ranking, decision to fix or not, management approval signoffs, actions taken, actual cost of fix, and the resulting effect on usability in further testing with users. Comparison of usability data with usability specifications and with usability data of previous iterations is crucial for managing the usability engineering life cycle for a given project, determining convergence of usability attributes, and (especially) deciding when to stop iteration.

CHAPTER 2. LITERATURE REVIEW

USER ACTION FRAMEWORK -DEVELOPMENTAL ASPECTS

Need for a Framework

There has been an increasing awareness of the importance of usability and because of this, organizations are expending ever increasing resources for "doing usability" – building enviable usability engineering laboratories, training developers in usability engineering methods, and conducting usability evaluations (Hix and Hartson, 1993). Though usability practitioners have effective methods to guide them in collecting raw observational usability data, due to a lack of a structural organization for the information collected and for the activity log, they have not obtained formidable success in usability development efforts. A review of many of the usability problems from real-world usability labs (e.g. Keenan, 1996) has shown that the reports recorded by the evaluators are often inconsistent, vague and incomplete. These reports are often committed to memory to be communicated to the designers who would fix the problems at a later time. This leads to poor quality problem reporting, which in turn leads to loss of information and misinterpretation of reports, eventually affecting the iterative usability development efforts. This poor communication of information can be directly attributed to a lack of a proper framework for analyzing / reporting usability problems and development activities.

In addition, very few software tools exist that have been shown to be effective in classifying, analyzing and reporting usability problems. There is no community database listing the problems reported by industry-wide practitioners, which would help other evaluators in their projects. Also, the existing user interaction development activities are limited to Graphical User Interfaces (GUIs) and are often not applicable to new styles of interfaces such as those found in web-based applications, virtual environments and voice I/O applications.

Development History

The first step to the current User Action Framework was the Usability Problem Taxonomy (Keenan, 1996; Keenan et al.., 1999). It was based on the view that a usability problem can be thought to possess two dimensions – task and artifact (Carroll, Kellogg & Rosson, 1991). These two dimensions can further be divided into sub categories. The task dimension has two categories, task mapping and task classification, and the artifact dimension has three categories, visualness, language and manipulation. These dimensions resulted in a taxonomy consisting of four levels of problem types and one level of specific example of usability problems. A study conducted to evaluate the reliability of this taxonomy (Keenan, 1996) showed that it yielded acceptable reliability on the artifact dimension but only marginal reliability on the task dimension thus concluding a better reliability is needed if this was to be used as a framework.

van Rens (1997) expanded the work started by Keenan by creating the Usability Problem Classifier, with the addition of new content and adjustment of structure. The most salient feature of this classifier was the classification of a problem relative to the timing of the problem as occurring before, after or during the user action. A "peel-off" mechanism was also created to rule out several less common but troublesome issues.

In a subsequent version, Andre et al. (1999) moved the "before, during or after" decision earlier in the classification process so that more of the task-based context will be at the beginning stages of classifying a problem. This ensured simplification of problem description and classification, but did not necessarily eliminate disagreements about classification results among usability practitioners. Other usability engineering support tools, like the usability inspector and usability data manager were also developed (Andre et al.., 1999) at around the same time when this version of classifier was developed. An efficient structure was needed for proper usage of the tools, to organize the usability concepts and issues in context of the tool. It was observed that in order to organize usability concepts and issues in the context of the purpose of the tool, each tool needs a structure. Instead of developing that structure for each tool, Andre et al. (1999) developed a unified design based on a central structured model of usability concepts and issues, for all tools, which allowed consistent structure, content and standard usability language across all tools.

In sum, the practice of usability engineering was envisioned to benefit from the following:

- A reliable framework which facilitates usability problem classification and high-quality problem reporting
- An integrated set of tools to support interaction development activities from usability testing, including usability problem classification and usability data maintenance
- Tools, including usability inspection tools, that can adapt easily to new interaction styles beyond GUIs

At present, the User Action Framework has five categories at the top level and each of the categories with its subcategories has been described earlier. All the tools share the UAF and have a web-based interface using DHTML, and active server pages to access a relational database.

Conceptual Model of the User Action Framework

The conceptual model behind the User Action Framework was an extension of Norman's (1986) theory of action model. Several other researchers (e.g. Cuomo, 1994; Lim, Benbasat & Todd, 1996; Rizzo, Marchigiani & Andreadis, 1997; Garzotto, Matera & Paolini, 1998) have used Norman's model in several ways and found it helpful for communicating information about usability problems. Norman's model has seven stages as follows:

1. Establishing the goal.
2. Forming the intention.

3. Specifying the action sequence.
4. Executing the action.
5. Perceiving the system state.
6. Interpreting the state.
7. Evaluating the system state with respect to the goals and intentions.

These stages deal with the interaction between a user and almost any machine consequently the conceptual model in the User Action Framework is also highly applicable to many interaction styles even outside computer applications. The stages in Norman's model were adapted and extended as the Interaction Cycle in the User Action Framework. The areas in the interaction cycle correspond to the seven stages in Norman's action model. Like Norman's model, the interaction cycle is a picture of how interaction between a user and any machine happens in terms of cognitive and physical user actions. In addition to this, a dual view has also been added, which is the machine (system) view of the same interaction and extended the concept to include interaction initiated by the system, by the environment or by the interaction cycles of other collaborating users (Kaur, Maiden & Sutcliffe, 1999). The interaction cycle is a cycle of actions depicting interactions of users with machines and also the categories of the higher levels of the User Action Framework. The User Action Framework content is about interaction design and how it supports the user and task performance during interaction. Hence, Norman's action model, merged with the structural concepts of the framework helps to address various issues when users interact with a machine.

USABILITY ENGINEERING

It is imperative to know what Usability Engineering is and why it is important before we can understand the methods of evaluating the usability of any interface. Usability engineering is the technique of setting quantifiable usability specifications, assessing interface usability, performing cost benefit analyses and making decisions regarding iterative improvements to the interface (Bennett et al., 1984; Carroll & Rosson, 1985; Good, Spine, Whiteside & George, 1986; Whiteside, Bennett & Holtzblatt, 1988). Usability Engineering helps to introduce a common platform for the design team and minimizes subjective aspect of the interface development. Usability engineering can be thought of as an important part in user centered design and represents the techniques, processes and methods for designing usable products and more importantly the philosophy that places the user at the center of this design (Rubin, 1994)

Web Usability and Evaluation

With the advent of the Internet era, there has been an explosion in Web-based interfaces and in the number of Internet users. The influence of the Internet is so strong as to introduce terms into the software vocabulary like "web year" which is a span of three months. Given such a short period of time, the developer has only a few weeks to create a functioning Website that should be able to evolve into something better over a few months. The importance of the usability of web-based interfaces is now more important than ever. Designers should be able to design a Website that provides useful content and also a satisfying experience for the user in tasks such as finding information, viewing images, reading and understanding text and downloading information and ordering products which will attract the user to come back to the site again (Lee, 2000). Testing of these sites to get a measure of their performance also becomes very important and many methods, like user-based assessment and quantitative assessment have been developed for evaluating websites. WebPages can be broadly classified into three types based on their content - web page as document, web page as document with product elements and web page as product with document elements (Hinderer & Kanter, 1998). They can also be classified as informational sites, search sites and transactional sites based on their purpose. Testing of web pages for usability is now complex due to hybrid existence of documents and software content in web pages and should encompass testing of both the elements. Two methods have been used to test the usability of websites in the past - Heuristic evaluation and laboratory evaluation (Kanter & Rosenbaum, 1997). Heuristic evaluation is conducted by experts who try to "put on a user's hat" and use set user interface guidelines and their prior experience as benchmarks to classify problems and give solutions. As the number of experts conducting the test increases, the number of problems identified also increases but the flipside is that they still are not representatives of a perfect user population. For web pages that have been up and running and have a good usability team in place, iterative heuristic evaluation will prove better. The alternative to heuristic evaluation is laboratory evaluation where the participants represent the user class the interface is targeted for. This method proves very effective for web pages that are being tested for the first time. Other than these two types of tests, a web page also receives feedback from its users who face problems while using it and report it to the administrator (Hinderer & Kanter, 1998). These reports, however, are not very informative as they do not convey why the user visited the website and also do not indicate why the user met some difficulty in using the site.

Formative Usability Evaluation

Two stages of evaluation exist for testing the usability of software applications and web-based interfaces - formative evaluation, which occurs early and also continues during the development of the system, and summative evaluation, which occurs after the user interface is completed. The testing of web pages should be done during various stages of development so the errors can be corrected as and when they are encountered (Kanter & Rosenbaum, 1997). This is exactly the purpose of conducting a formative usability evaluation, which is defined as the evaluation of interaction design as it is being developed, early and continually throughout the interface development process (Carroll, Singley & Rosson, 1992). The developer can be asked to develop prototype (paper / low, high fidelity) of the interface during the early stage so that it can be tested for usability attributes like ease of use and navigational functionality. The test can be administered through a questionnaire (Kirakowski & Cierlik, 1998) to be filled out by the user after they have completed a given task or though extensive interrogations of the participants after completion of the assigned task. This will test the comprehensibility of page headings, links and button names and also the usefulness of the navigation panels. After the page is fully developed, it is tested again and the test focuses on content and the "look and feel" of the site while also considering navigation and ease of use.

The traditional methodology for conducting a test is derived from the classic approach for conducting a controlled experiment. The test methodology includes formulating a specific hypothesis, testing the variables involved by isolation and manipulation, and analyzing cause and effect relationship leading to either the confirmation or rejection of the hypothesis. However, this would not be a suitable approach for evaluating web pages in fast paced development environments due to various political and organizational constraints (Rubin, 1994). A test based on the classical approach to obtain valuable results requires:

- Experienced usability specialists with adequate prerequisite knowledge for interpreting the results which would otherwise be misleading
- Randomly chosen participants
- Large sample size

It is very difficult to meet all the above-mentioned requirements while conducting formative usability evaluation in fast paced development environments. A slight modification to the traditional methodology can be used in formative evaluation studies and is more suitable for real world situations (Prescott & Crichton, 1999) and the modified procedure is itemized below:

- Develop test objectives rather than hypotheses
- Use a participant group that is representative of the user class and not necessarily random
- Represent the actual work environment

- Observe the end users that use or review a product. Users may be asked to report their responses directly or may be observed by an evaluator to obtain their responses.
- Obtain qualitative and quantitative performance and preferences measures.
- Recommend improvements to the design.

Data Collection in Usability Evaluation

The Critical Incident method is a cost-effective method to capture data and the captured data is of high quality and can therefore be directly converted to usability problems. The data, in this evaluation, is collected by recording critical incidents of the user and hence this method is called the “Critical Incident” method. A Critical incident of a task is the most important information associated with task performance (Hartson, & Castillo, 1998). The information is gathered by the users themselves or by evaluators and is then used to analyze and formulate a set of usability problems and to find solutions to these problems. In case of user reported critical incident method, when a user working with an application encounters a difficulty (negative critical response) or accomplishes a given task (positive critical response), the response (incident) is conveyed to the evaluators by the user through a contextualized critical incident report. The success of this method depends on the capability of the users to identify and report critical incidents directly. Flanagan (1954) employed trained observes (not users), who were domain knowledgeable data collectors, making observations of ongoing activities in the user’s normal working environment. The critical incident technique, as followed by most studies these days, has been reformed and adapted for human computer interaction in such a way that even untrained users can identify critical incidents (del Glado, Williges, Williges & Nixon, 1986). In case of an evaluator interrogating the user to record the critical incidents, the evaluator directly observes the user and produces a list of critical incidents of the user during task performance. This information is later conveyed to the designers and is used for correcting any problems that may have occurred.

RELIABILITY STUDY

Importance of Reliability

Reliability studies of a software application or product are very important to ascertain that the outcome of tasks performed on the product is independent of the user using the product. If a product with low reliability is used for a particular purpose, the results obtained will vary from practitioner to practitioner and will not provide quality data for the project. An example of this is the Heuristic Evaluation method developed by Nielson and Molich (1990). Studies conducted on heuristic evaluation (Doubleday, Ryan, Springett & Sutcliffe 1997; Jeffries, Miller, Wharton & Uyeda, 1991) have shown that though heuristic evaluation was considered to be a cheap and easy method, it did not provide enough distinction between various problems (Dutt, Johnson & Johnson, 1994). The reason may be that heuristics are often too general for detailed analysis with each heuristic covering a broad range of factors, resulting in overlaps in categories and gaps (Sears 1997). This leads to mis-classification and hence unreliable results. Hence, it is important that the any too, like the User Action Framework, has good reliability for the usability professionals to use the tool as a common foundation for various activities. Although the reliability of a method appears to be inherently important, the literature on reliability studies on usability methods is essentially non-existent.

Reliability Measurement

Many methods are available to measure reliability (Meister, 1985). A study by Andre et al. used the kappa statistic (Cohen, 1960) to measure the level of agreement between the users of the User Action Framework. Chance agreement is measured and expressed as a value ranging from –1 to +1, with –1 denoting less than chance agreement, 0 denoting only chance agreement and +1 denoting more than chance agreement.

Reliability of the User Action Framework

A reliability study of the User Problem Taxonomy conducted by Keenan (1996) showed that the classification of the User Problem Taxonomy was reliable at the primary level ($k = 0.403$, $p < 0.001$) but not reliable at the task dimension ($k = 0.095$, $p > 0.1$). van Rens (1997) expanded the work started by Keenan by creating the Usability Problem Classifier, with the addition of new content and adjustment of structure. This conceptually simplified problem description and classification, but did not eliminate disagreements about classification results among users.

The User Action Framework was an extension of the User Problem Taxonomy developed by Keenan (1996) and the Usability Problem Classifier developed by van Rens (1997). It was a structured framework that was developed based on Norman's action model (1986) with an aim to provide better reliability for classifying and analyzing usability problems. Norman's action model dealt with interactions between human and any machine and hence a model based on this would be suitable for any kind of interface.

The results of the reliability study conducted by Andre et al. (2000) showed that the User Action Framework attained a better reliability ($k = 0.583$, $p < 0.001$) with regard to the overall agreement than the User Problem Taxonomy. The degree of user agreement at different levels is also greater than the agreement at the highest level of the Taxonomy.

In the latest version of the User Action Framework, some changes have been made in terms of content and its quasi-hierarchical structure. Also, the two sub-categories under Planning (High-level planning and Translation) were eliminated and translation was added as a separate category. Due to these changes incorporated into the framework, it was necessary to study the reliability of the User Action Framework and hence the intention is to conduct the reliability study in the manner as followed by Andre (1999).

CHAPTER 3. FORMATIVE USABILITY EVALUATION

The main purpose of the formative usability evaluation was to assess the usability of the web interface of the UAF Explorer and the presentation of its content by gauging its ease of use, and understandability. The focus is only on the UAF Explorer, since it is considered a gateway to explore and learn the content of the User Action Framework and all the other components of the tool set are still in development phase. As mentioned earlier, the goal was to uncover as many usability problems as possible and fix them, so that usability will not be a confounding factor in the reliability study.

METHOD

The formative usability evaluation of the UAF was done in two iterations. Three participants took part in the first iteration and four participants in the second iteration. At the end of the first round of studies, the interface was re-designed based on the results before continuing with the second round.

Formative Evaluation of the UAF – 1st Iteration

Participants

Three participants were selected for the first round of studies. Since the target users of the User Action Framework are both experienced and inexperienced usability practitioners, it was decided (R.H. Hartson, May 2001, Personal Communication) that the participants would fit two user profiles – people who had been exposed to usability engineering activities (at least one year of usability engineering activities) and people who were new to the usability-engineering field. For the first iteration, all the three participants had a Masters degree in Computer Science and had at least one-year of experience in user interface design, usability testing and/or evaluation. The participants worked in companies in and around Blacksburg, Virginia.

Materials and Testing Facilities

Materials for this formative evaluation study included a working high-fidelity prototype of a website containing the User Action Framework content loaded on a computer and available for use by the participants. Though the website contained all of the components of the suite of tools, the participants primarily used the UAF Explorer tool. Figure 3-1 shows the UAF Explorer start page with five areas of the interaction cycle (Planning, Translation, Physical Actions, Assessment and Independent).

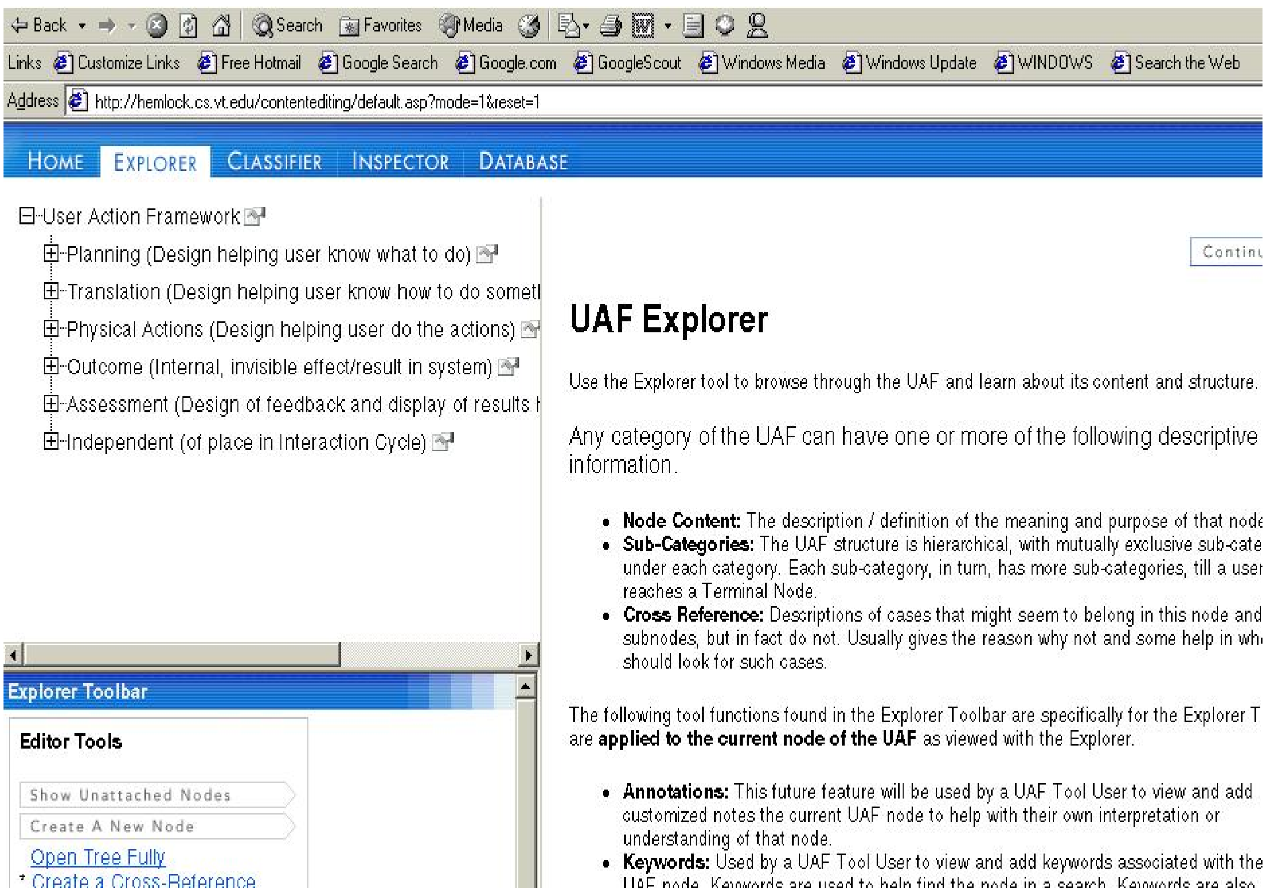


Figure 3-1: Screen shot of UAF Explorer tool

The formative usability evaluation study took place at the Usability Methods Research Laboratory at Virginia Tech. The lab consists of a user room, equipped with a Pentium-based, Windows 2000 workstation on which the web site for User Action Framework will be loaded. A videocassette recorder will be used to record both audio and video of the test sessions. A digital recorder software was used to record participants' path through the website. The software also allowed mixing audio with the screen images and storing the file onto the computer. In all the test sessions, the test monitor sat beside the participant to encourage him/her to talk aloud and prompt for feedback. A schematic of this is shown in Figure 3-2.

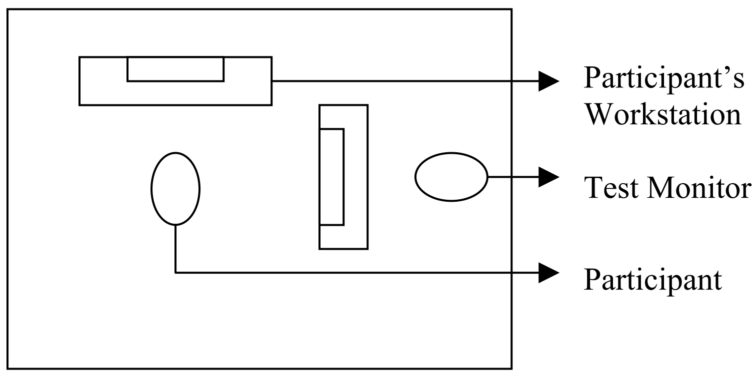


Figure 3-2: Schematic of Observation Setup

Procedure

The participants were greeted by the test monitor and made to feel comfortable and relaxed. The participants were asked to read a description of the purpose of the study and were asked to sign the Informed Consent Form (Appendix B). Prior to data collection, the participants received a short, verbal, scripted introduction and orientation to the test, explaining the purpose and objective of the test. Then the participants began the session with a tour of the Usability Methods Research laboratory followed by an explanation of how the data would be collected during the study. At this time they completed a pre-test questionnaire (Appendix C) that gathered basic demographic information. For the first iteration, the summary of data from the pre-test questionnaire has been given in Table 3-1. Participants were then encouraged to explore the UAF interface while verbalizing their thoughts through thinking aloud. The participants were provided with a task list with instructions, containing scenarios relating the functionalities of the UAF Explorer and some classification tasks (Appendix D). The task scenarios were selected from a database, which contains about a hundred such usability scenarios. The participants were asked to start reading the task list and to perform the tasks in a way that is typical and comfortable to them. The participants were allowed to learn from exploring various paths until they reached the end state. An identification number was created for each participant, which was used to correlate data with a particular participant. Breaks were provided between the tasks if requested by the participants.

Table 3-1: Pre-test questionnaire data from 1st iteration

Category	Participant #1	Participant #2	Participant #3
Age (range)	39-48	18-28	29-38
Gender	Female	Male	Male
Completed education	Masters – Computer Science	Masters – Computer Science	Masters – Computer Science
Computer usage (years)	13	18	15
Computer platform	IBM-PC, Mac	IBM-PC	IBM-PC
Operating system	Windows 9X, Mac OS	Windows 9X/00,	Windows NT
Internet/Web usage (years)	10	7	8
Web Browser	Internet Explorer, Netscape Navigator	Internet Explorer,	Netscape Navigator
Proficiency in using the Web (out of 10)	8	8	9
Experience in usability engineering activities (years)	1-2	3-4	1-2

The test monitor sat in the room with each participant while conducting the test. The test monitor initiated tasks after orientation, made notes about relevant participant behavior, comments and any unusual circumstances that might affect the performance of the participants. The participants were encouraged by the test monitor to verbalize their thoughts by thinking aloud while performing the task. The test monitor will make sure that no help was provided unless a question about the test procedure arises. Care was also taken to enable the participants to do the tasks but not lead them by unintentionally providing hints about correct performance. Finally, the test monitor debriefed the participants and thanked them for their participation in this study.

Data Collection

The data collected in this study was observational data on task performance and verbal protocol data as provided by the participants. The observational data included critical incidents as observed by the test monitor, any other comments / responses the participants made throughout the test session and comments at the end of test session. The observational and verbal protocol data were reviewed and edited by the test monitor, based on retrospective viewing of the videotapes. Table 3-2 summarizes the data collected during the first round of studies categorized under two categories – positive critical incidents and negative critical incidents. These are the set of incidents used to identify usability problems, based on which re-design recommendations were made. A positive critical incident is defined as an incident, which signifies an increase in speed or accuracy of task performance (Thompson, 1999). Rubin (1994) defines a negative critical incident as one, which disrupts the performance of users. A consolidated list of all the critical incidents and user comments for the both the iterations were given in Appendix E.

Table 3-2: Summary of data for 1st iteration

Negative Critical Incidents	Positive Critical Incidents
1. User unable to locate his position in the UAF. 2. Clicked on a cross-reference. Commented that this is one of the sub-category choices 3. Confused at the terminal node page. Did not know what to do. Some thought they had to continue with the classification 4. Confused by the word “Object” in various places 5. Error Messages Vs. Information Display under Assessment. Did not understand the difference 6. Users did not understand the difference between the words – “Tool Button” and “Tabs” in the home page 7. User commented that the difference between the first two categories in Content, Meaning under Translation was not clear. Subsequently, the user got the classification wrong. 8. Users were not able to match the terms in the home page with the content inside the tool 9. Did not identify some links in the UAF website 10. Got confused when clicked on “Upto Parent”. Users expected to be taken to previous sub-categories, but in fact, were not. 11. User commented that the classification of usability problems should be specific to a particular application 12. Two users expressed concerns about high number of sub-category choices in a level in the UAF	1. Noticed the navigation bar at the top of the UAF web pages 2. Understood what “Physical Actions” means using the description 3. Mapping of words in the UAF with particular scenarios, to aid classification 4. Identified pluses and minuses in the top left-hand frame as a tool for rapid navigation in the UAF 5. Users commented that this tool can be used for as a common framework

Results – 1st Iteration

Usability Problems and Re-design Solutions

This section lists the usability problems identified through observation of task performance and from participants' comments during the test sessions. Re-design solutions were provided for each of these usability problems. These re-design solutions were arrived at by means of brain storming sessions involving the author and Dr. Rex Hartson (H. R. Hartson, October-2001, Personal Communication). During these sessions, the negative critical incidents, presented in Table 3-2, were analyzed and a list of usability problems was produced. Each of these usability problems was then studied, and re-design decision was taken specific to that particular problem. Expert judgment was primarily used while deciding on re-design solutions; this ad-hoc procedure matches the standard industry format of "converting" usability data to specific design elements. There is a lack of a standard methodology and associated literature, which talks about transforming usability data collected during evaluation studies to global and specific design components. It is expected that the User Action Framework will be used as a common platform to do this "conversion". Also, some screen shots of re-designed pages have been provided, which reflect the re-designed web pages.

Problem – User losing track of where he/she is: The users lost track of where they are in spite of the classification path given at the top of each page.

Re-design Solution: User "locations" were clearly indicated using a Navigational Bar at the top of each page. The Classification Path they followed was clearly indicated and the navigation bar was visually differentiated from the rest of the page (Figure 3-3).

Figure 3-3: Re-designed web page with the Navigational Bar

Problem – Confusing UAF Content: There were few instances where the users had problems with the Content of the User Action Framework. Table 3-2 presents such instances and the changes that were made had been explained.

Table 3-3: Changes made to the Content of the UAF

Before – Problems with Content	After – Re-design Solutions
Users confused by the word "Object" in various places in the UAF. Several users thought "Object" means some piece of code.	To change the word to "User Interface Object", so the confusion that "Object" means some piece of code was eliminated
Confusing wording in sub-categories for Assessment – "Error Messages" Vs. "Information display"	To change the wording of choices to "Issues about Feedback" and "Issues about Information Display". Also, to add more examples to each sub-category.
Inconsistencies when using the words "Tool Button" and "Tabs" in the Home Page	To use those words consistently in the Home Page and also across all the pages in the UAF Explorer
Lack of clear distinction between the explanations for the first two sub-categories in Content, Meaning under Translation	To make the distinction more clear my adding more explanations to those sub-categories. Also, to add examples to each sub-category

Problem – No clear distinction between Sub-categories and Cross-references: Users were often confused between the sub-categories of a particular component of the interaction cycle (e.g.: Planning) and the cross-references, which were presented to the users in some pages.

Re-design Solution: The two sections of the page were visually differentiated using different colors. It was clearly said that the cross-references are not a part of sub-categories, rather links to other parts in the UAF (Figure 3-4).

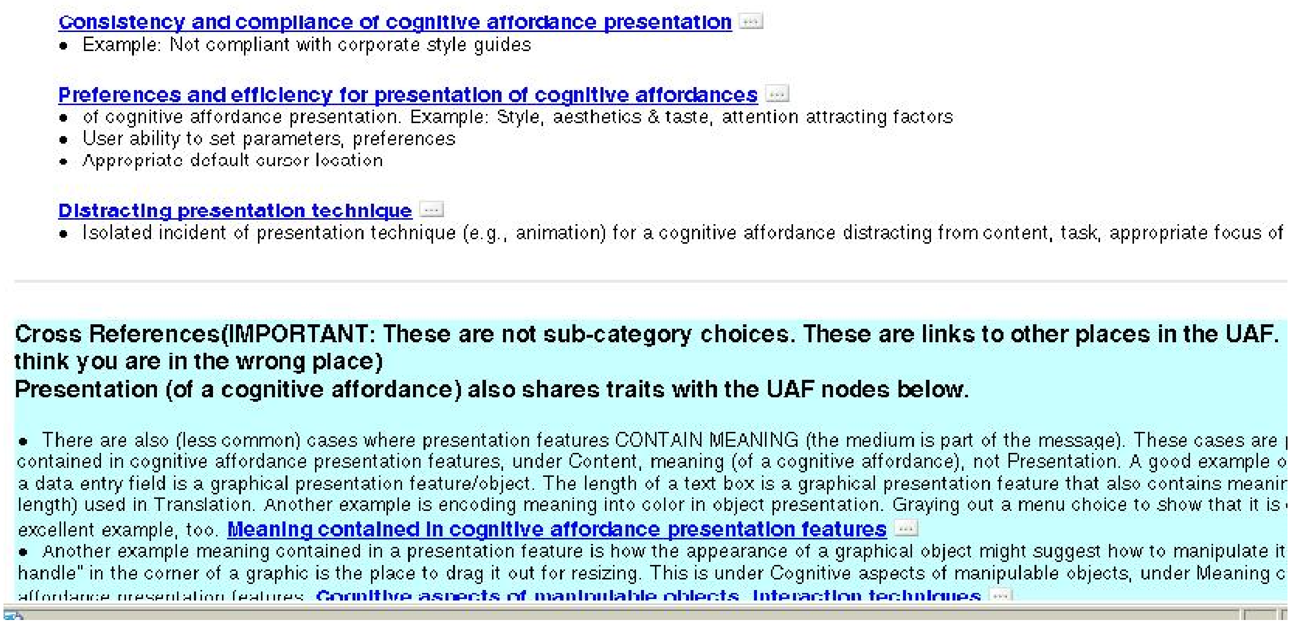


Figure 3-4: Re-designed Cross-references page

Problem – No indication for the terminal node: There was no indication when the users reached the Terminal Node. They often thought they had to continue with the classification, somehow.

Re-design Solution: Users were indicated clearly in the navigation bar that they have reached the Terminal Node (Figure 3-5).

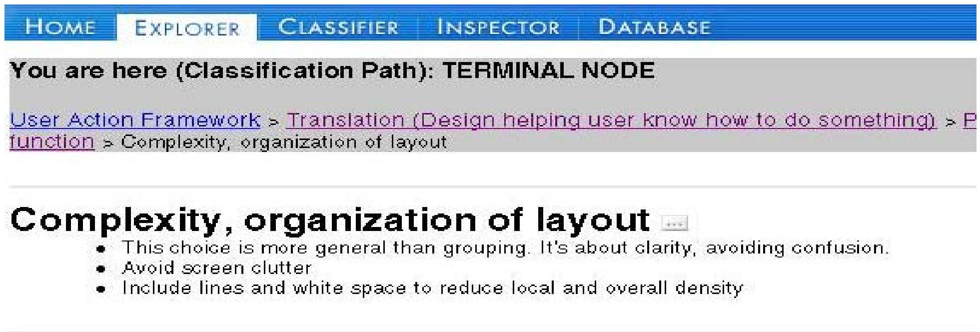


Figure 3-5: Re-designed Terminal Node page

Problem – Links not being underlined: Some users had problems with the fact that the links were not underlined in the UAF Explorer web pages.

Re-design solution: All the links were underlined to match the users' mental model about the links on the World Wide Web and also to maintain consistency.

Problem – Outdated home page: The home page did not exactly reflect the updates that were made in the UAF Explorer before this round of formative evaluation. Users were confused because the content / terminologies in the home page did not match the same in the UAF Explorer.

Re-design solution: Home page was updated to exactly reflect the content / terminology across all pages in the UAF Explorer. It was also mentioned that different tools exist by using bold font face (Figure 3-6).

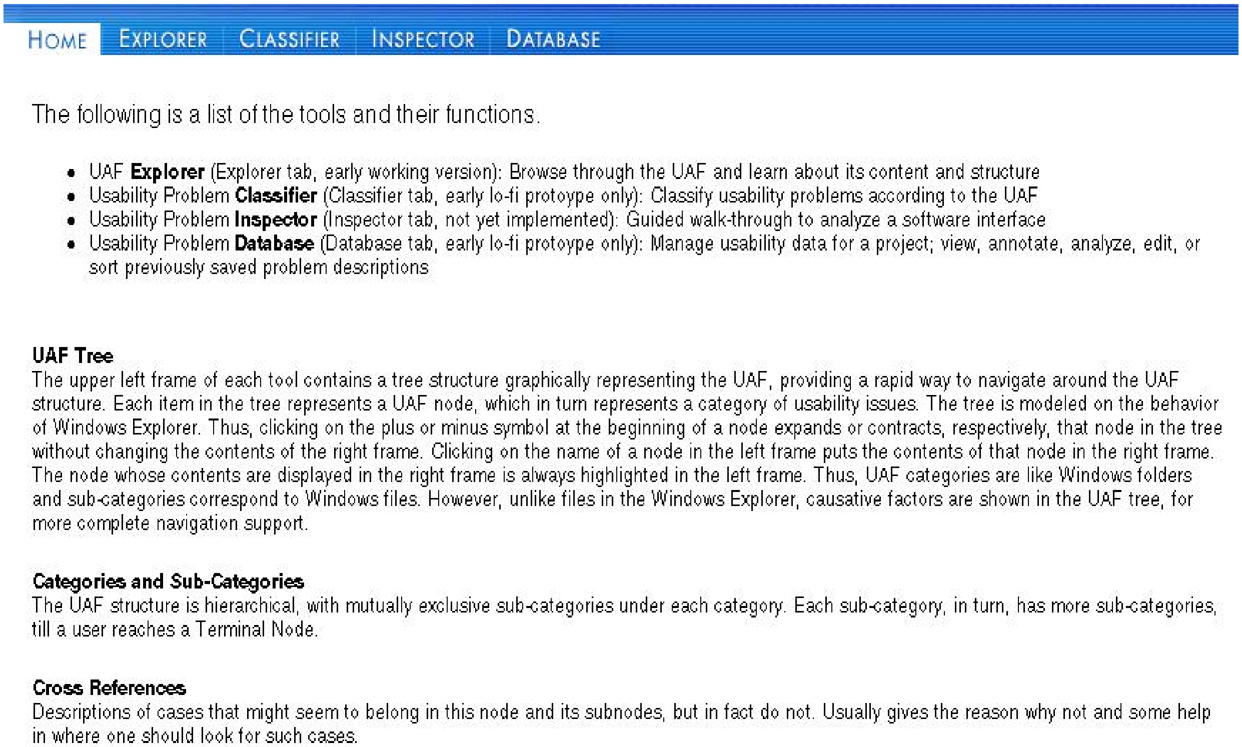


Figure 3-6: Updated Home Page

Problem – Bug in the code: The button - Up to Parent did not work the way it was supposed to. It was supposed to take the user to the previous level in the hierarchy of the User Action Framework.

Re-design Solution: The code, which fired this action, was fixed. When users clicked on the button – Up to Parent, they were taken to the previous level in the hierarchy of the User Action Framework.

Formative Evaluation of the UAF – 2nd Iteration

The testing facilities, procedure and the data collection methodology followed in the 2nd iteration were the same as in the 1st iteration, and hence they were not described in this section. A brief explanation was made about the participants for this round of studies and Table 3-4 features the summary of data from the pre-test questionnaire. The materials for the 2nd iteration included the web-based UAF Explorer tool, but this time only the re-designed website was used. As it can be seen in Table 3-5, the negative critical incidents in the second iteration got considerably reduced and hence fewer number of usability problems were identified as a result. It should also be noted that the positive critical incidents reflect the changes made at the end of first round of studies.

Participants

Four participants were selected for the second round of studies. All the four participants were selected from students enrolled in CS 5714: Usability Engineering (Fall, 2001), majoring in Computer Science, Human Factors, Psychology or Industrial Engineering. The participants had, at least, a bachelor's degree in Computer Science, Human Factors, Psychology or Industrial Engineering.

Table 3-4: Pre-test questionnaire data from 2nd iteration

Category	Participant #1	Participant #2	Participant #3	Participant #4
Age (range)	18-28	18-28	18-28	18-28
Gender	Female	Female	Male	Male
Completed education	Masters – Computer Science	Bachelors– Computer Science	Bachelors– Industrial Engineering	Bachelors– Psychology
Computer usage (years)	5	8	4	6
Computer platform	IBM-PC	IBM-PC	IBM-PC	IBM-PC, Mac
Operating system	Windows 00	Windows 00	Windows 00	Windows 00, Mac OS
Internet/Web usage (years)	4	4	4	6
Web Browser	Internet Explorer	Internet Explorer	Internet Explorer	Internet Explorer
Proficiency in using the Web (out of 10)	5	6	8	8

Table 3-5: Summary of data for 2^{nd} iteration

Negative Critical Incidents	Positive Critical Incidents
1. Participants did not know how to start a new classification, even if they knew they had reached the terminal node. 2. One user did not see the "Continue" button at the bottom a page. 3. One user commented the need for more rapid navigation techniques to browse the UAF web pages 4. Users commented about the need for more examples under sub-category choices	1. Users were able to distinguish between sub-category choices and cross-references 2. Users commented that Navigation Bar was clearly identifiable and used the "You are here" to locate their position. They also used the links in the navigation bar to jump to previous sub-categories in the UAF 3. Terminal Node was clearly identified using the indication at the Terminal Node page 4. Users selected the correct sub-category under Assessment. 5. Participants used the home page to identify the content inside the tool. 6. Mapping of words in the UAF with particular scenarios, to aid classification 7. Identified pluses and minuses in the top left-hand frame as a tool for rapid navigation in the UAF 8. Users had no problems identifying the links in the UAF web pages 9. Participants used "Upto the Parent" button to go to previous sub-category choices

Results – 2^{nd} Iteration

Usability Problems and Re-design Solutions

This section lists the usability problems identified through observation of task performance and from participants' comments during the 2^{nd} round of formative evaluation studies. Re-design solutions have also been given, which are arrived at using the same methodology used for the 1^{st} round of studies – brainstorming sessions involving the author and Dr. Rex Hartson (H. R. Hartson, October-2001, Personal Communication).

Problem – No clear indication of how to start a new classification: When the users reached the terminal node, they were confused as how to start a new classification.

Re-design Solution: A button was added in the terminal node page, with label – " Go to Top of the UAF". This way, the users would know they had to click on the button to start a new classification (Figure 3-7).

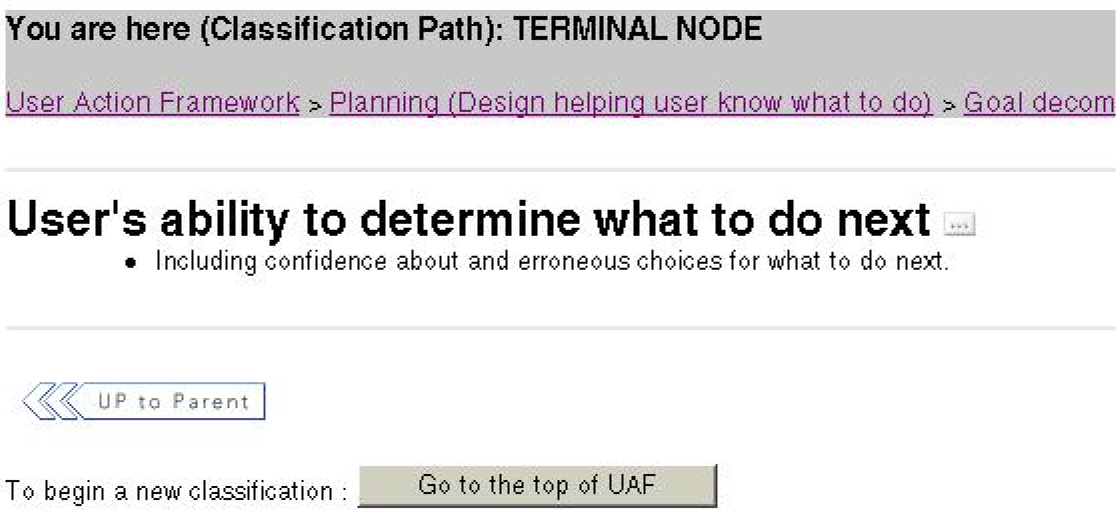


Figure 3-7: Button at the Terminal Node page

Problem – "Continue" button in Explorer Start Page: The users had to scroll down to see the "Continue" button, since the button was placed only at the bottom of the page. The users could not see this if they did not scroll down the page.

Re-design Solution: "Continue" button was duplicated and placed on the top of the page, also. This way, the users that did not scroll had a chance to see the button at the top of the page (Figure 3-8).

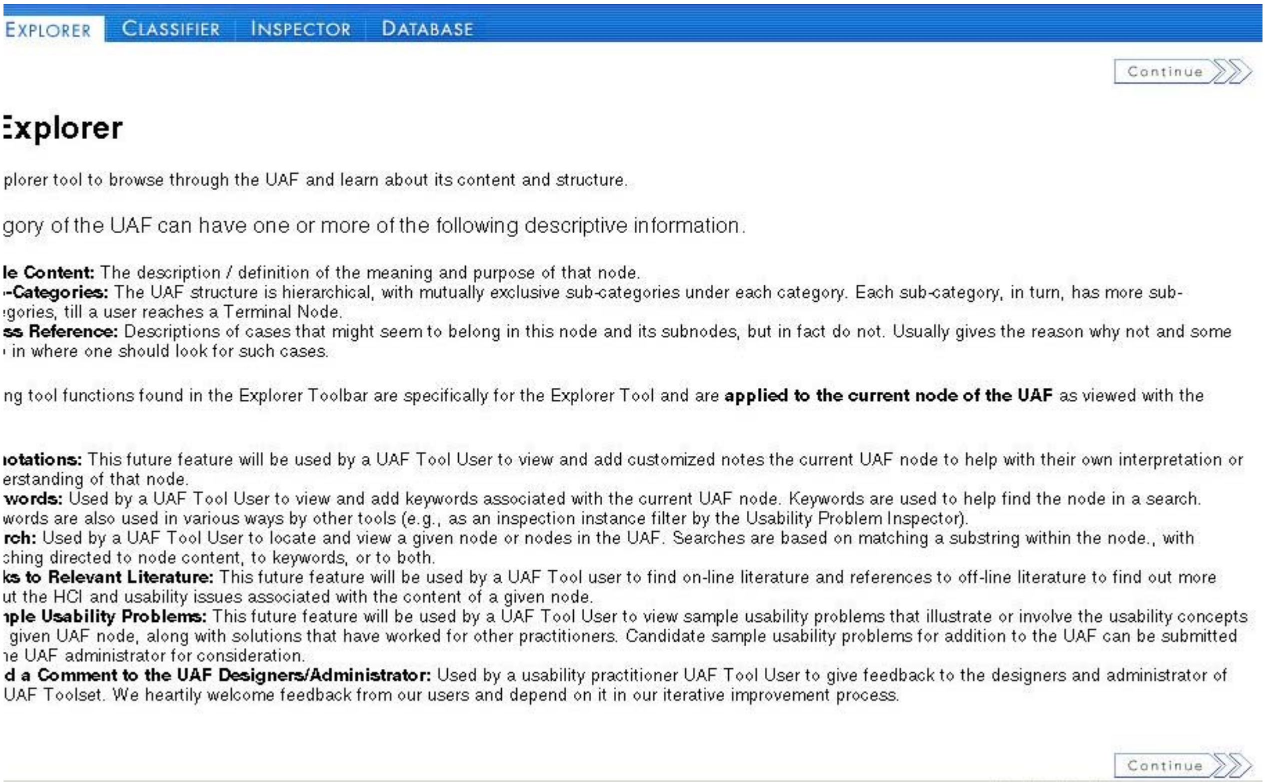


Figure 3-8: Re-designed Explorer Start Page

DISCUSSION

Overall results from this study showed that the UAF was an effective and usable tool for usability problem classification. The critical incidents data and users' comments to support these claims were presented earlier in this chapter. The users reported that the use of colors to visually differentiate between sub-categories and cross-references was, indeed, useful to distinguish between two parts of a page. The participants acknowledged the navigation bar, presented at the top of the page, as a "position locator" and to jump to previous sub-categories when classifying a usability scenario, without using the back button in the web browser. Also, The participants identified the Windows Explorer paradigm in the upper left-hand side of the Explorer interface as a very useful tool. The pluses and minuses (Figure 3-9) in the interaction cycle tree were viewed as a rapid way to navigate around the User Action Framework and to identify specific parts of the interaction cycle without browsing through the entire website.

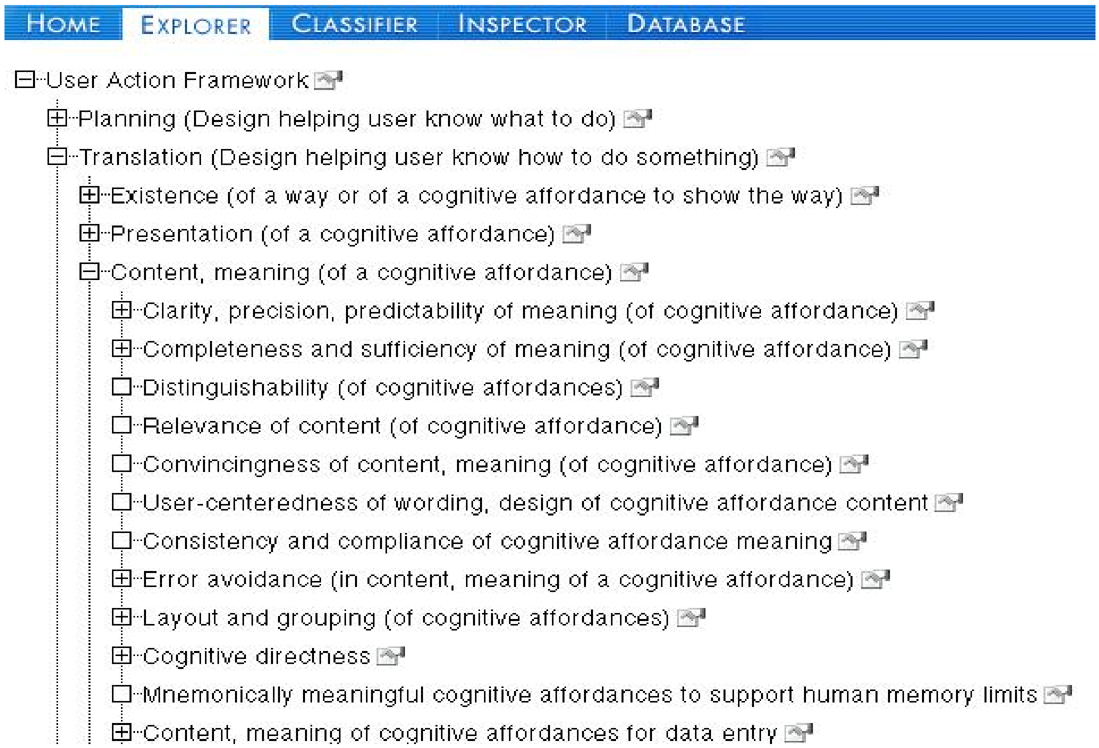


Figure 3-9: Enlarged view of the Windows Explorer paradigm

During the first round of iteration, most of the users were able to obtain the correct classification path, although they did not classify it correctly the first time. However, they had problems traversing the UAF Explorer website. The problems were identified using the critical incidents data and user comments collected during the study. Detailed explanations of those problems and re-design solutions aimed at rectifying those problems were presented in earlier sections of this chapter. The re-designed website was used during second round of formative usability evaluation; the users, during second round of iteration, felt that the usability of the interface got better in terms of navigation, ease of use and identification of specific areas of the interaction cycle. The usability problems identified got considerably reduced; design solutions were arrived at and an effort was made to fix those problems, as well.

According to Kanter & Rosenbaum (1997), usability testing of web pages should be done during various stages of development so the errors can be corrected as and when they are encountered. The purpose of performing the lab-based formative usability test was exactly this - to determine, at an early stage, if the theory-based framework and tool could be effectively used to classify usability scenarios according to the hierarchical structure of the UAF. Waiting until late in the development process when much of the interface has already been implemented, it will be far more difficult to make changes indicated by usability evaluation (Hix & Hartson, 1993). The lab-based testing generated a set of usability problems known to affect the performance of the users. Those problems addressed the usability of the UAF Explorer in terms of its interface and content. Before proceeding with the reliability study, the problems were fixed by means of a re-design effort, lead by a set of recommendations.

This work produced an improved interface of the UAF Explorer in terms of its ease of use and navigability. It is the hope of the author that this piece of research would provide usability professionals with a useful and usable tool to conduct various usability engineering support activities.

CHAPTER 4. RELIABILITY STUDY

The goal of the reliability study of the UAF was to allow experts to use the UAF for classifying certain usability problem descriptions and to record how well they agree with one another.

METHOD

Participants

Ten participants were recruited for this study. The participants included professionals working in commercial and educational institutions where usability testing/analysis is a part of their job description. All participants had at least a bachelor's degree in Computer Science, Human Factors, psychology or industrial engineering and had sufficient experience (at least 1year) in user interface design, usability testing and/or evaluation. Care was taken not to include the same participants used by Andre (1999) in his reliability study or the participants used in the formative evaluation study of this work.

Materials

Materials for the reliability study included a working high-fidelity prototype of a website containing the User Action Framework content loaded on a computer and available for use by the participants. Though the website contained all of the components of the suite of tools, the subjects primarily used the UAF Explorer tool to classify the usability problems. Figure 4-1 shows the UAF Explorer start page with different areas of the interaction cycle. By selecting one of the interaction cycles (e.g. Translation) participants started traversing the structure as shown in Figure 4-2.

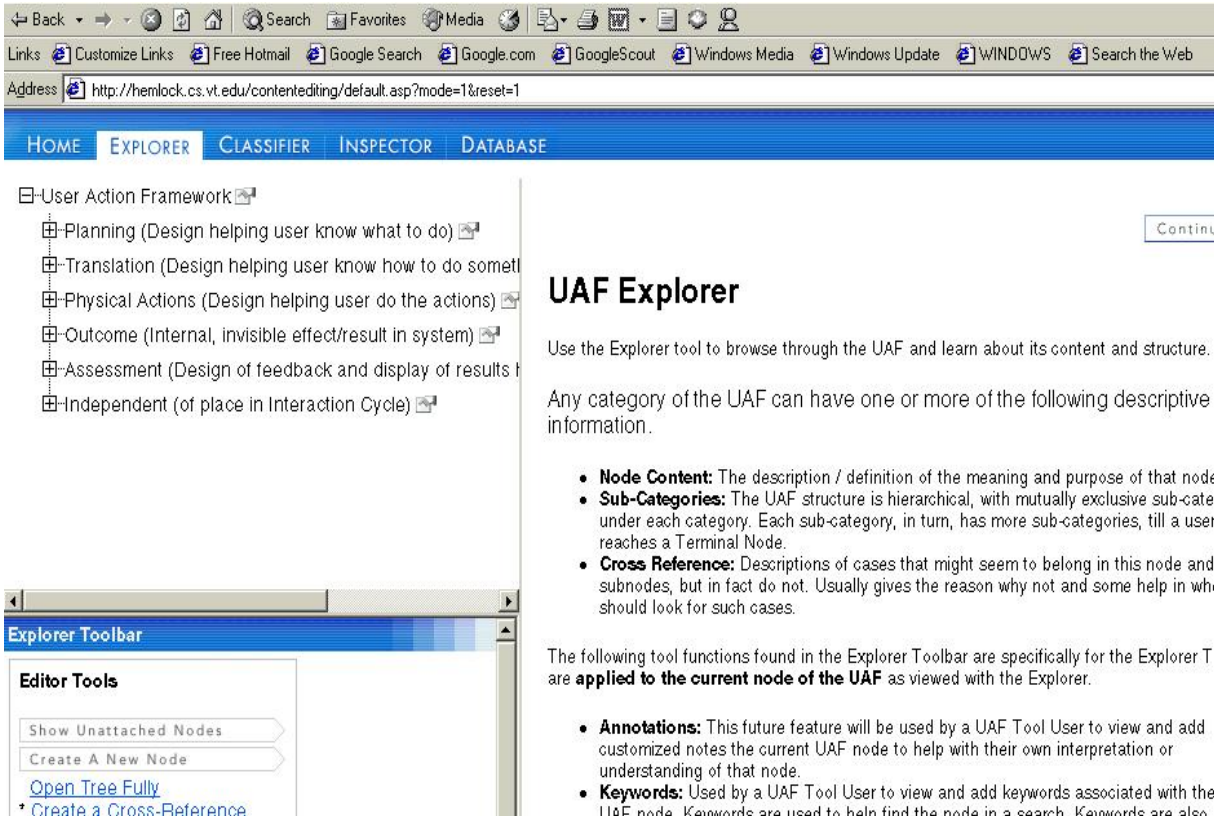


Figure 4-1: Start page for the UAF Explorer

Fifteen usability problem case descriptions were chosen from a database, which contains about a hundred such usability problems. The problems were selected based on their real world expected frequency of occurrence. The majority of usability problems were found to fall under translation followed by assessment, physical actions and planning (Hartson et al., 1999). Scenarios have been developed based on these usability problems and the participants were provided with the scenarios and instructions to classify (Appendix G) rather than a simple statement of a particular usability problem, as given in the Table 4-1.

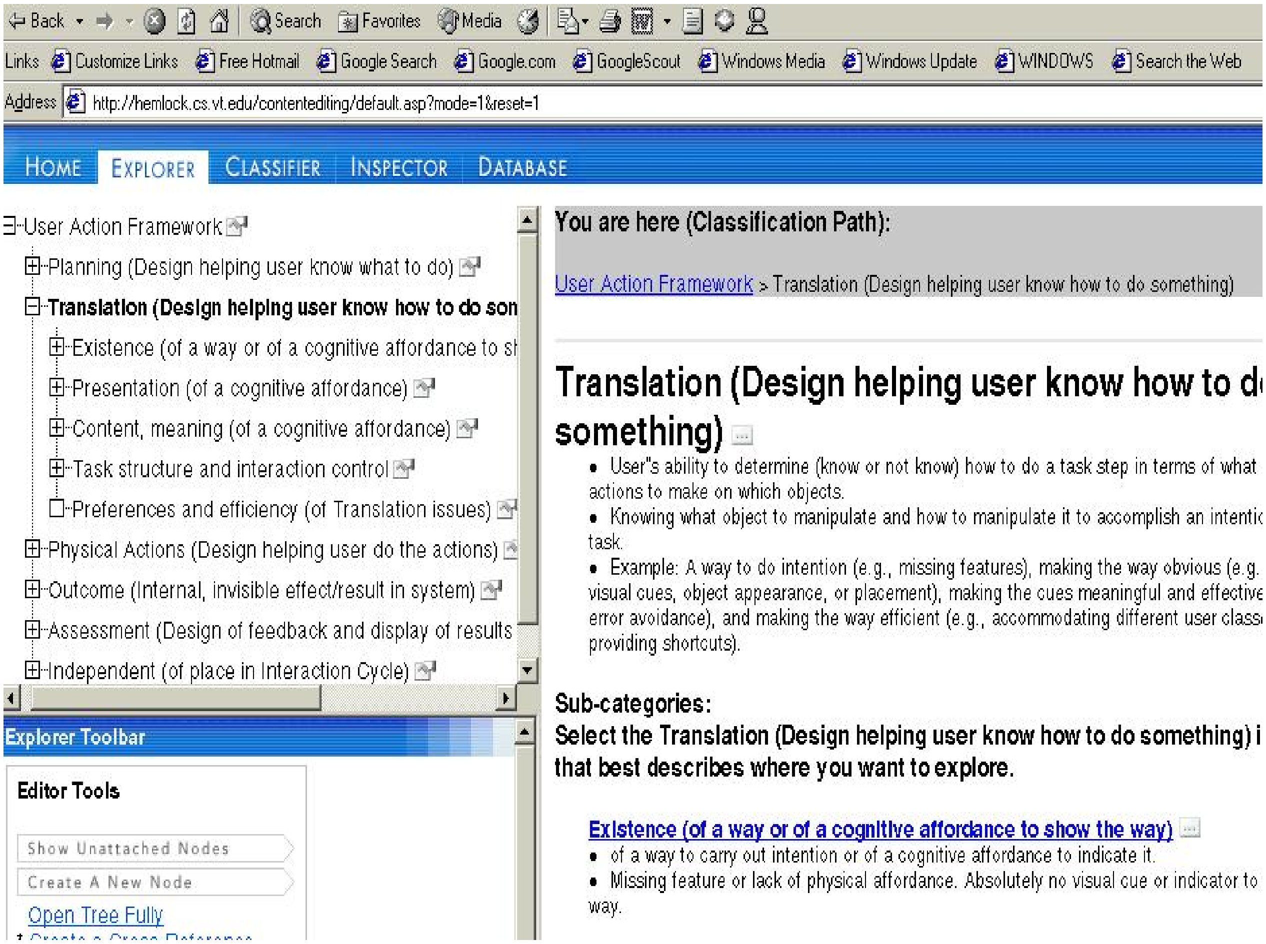


Figure 4-2: Translation page in the User Action Framework

Table 4-1: Usability problems for reliability study

Problem no.	**Type of usability problem**	**Relevant area in User Action Framework**
1	Unreadable error message	Assessment
2	User does not understand master document feature	Planning
3	User cannot find a feature to support re-using document numbers in a document retrieval system	Translation
4	User clicks on wrong button	Physical Action
5	User cannot directly change a file name in an FTP program	Translation
6	User cannot tell if system is performing requested operation	Assessment
7	User wants to fix database error but is confusing by button labels for appropriate action	Translation
8	Program does not provide a Ctrl-P shortcut for printing	Translation
9	User cannot understand the error message provided by system	Assessment
10	Unnecessarily long error message	Assessment
11	User cannot distinguish a button because of background	Physical action
12	Data does not see way to select odd font size	Translation
13	Data entry format not provided	Translation
14	Uncontrollable scrolling	Physical action
15	Vision impaired user needs preference options for setting larger font size	Translation

Procedure

The reliability study was conducted at the Assessment and Cognitive Ergonomics Laboratory at Industrial and Systems Engineering department in Virginia Tech. The lab consisted of a user room, equipped with a Pentium-based, Windows 2000 workstation on which the web site for the User Action Framework was loaded. Before beginning the actual evaluation, each user was asked to read and sign the informed consent form (Appendix H). They were then instructed to go through an on-line training program about the UAF and how the UAF Explorer can be used to classify the usability problems. This training covered the structure of the UAF, components of the interaction cycle and a walk-through of the classification process using guided classification examples. After the training, the participants were asked to classify six example problem descriptions according to the structure of the UAF. It was required that each participant should start with the correct first level of classification for at least four out of the six problem descriptions. Otherwise, the participant was to be excused and was no longer required to continue with the study. As it turned out, all the ten participants started with the correct first level of classification, and hence were allowed to proceed with the actual evaluation tasks.

The participants read the case descriptions (scenarios) given to them and then started to classify the usability problems. The users could go over any number of paths in the User Action Framework before reaching the final classification of the usability problem. However, users were asked to classify all possible causes of the usability problem in the interaction design. Since the primary focus of the study is proper classification of the usability problems, or at least agreement about classifications among the users, the time to complete the classification was not controlled.

Hypothesis

The hypothesis in this reliability study is that the UAF will result in an overall reliability score that is significantly greater than that of the overall reliability score obtained by Andre (1999). Previous research by Andre (1999) resulted in a strong overall agreement ($k = 0.583$, $p < 0.001$), indicating agreement is greater than would be expected by chance.

Data Collection and Analysis

The most commonly used reliability measure is the kappa statistic (Cohen, 1960) for categorical lists especially in cases where there is a possibility of agreement between the evaluators. The present study uses more than two evaluators, and hence requires the extension to the kappa statistic provided by Fleiss (1971), as kappa has traditionally been used only for comparison between two observers. The value of kappa is an indication of the extent of agreement between the evaluators. It is scaled between +1 and –1 with positive values of kappa indicating greater than chance agreement, 0 indicates only a chance agreement, and negative values of kappa indicates less than chance agreement. Since kappa is normally distributed, it can also be used to test whether agreement exists beyond the chance agreement level. Kappa for this study, denoted by k_1, is calculated as follows:

$$k_1 = (Po\text{-}Pc) / (1 - Pc)$$

where, *P*o is the proportion of observed agreement and *P*c is the proportion of agreement expected by chance. The formula and description of how to calculate Po and Pc are given as follows:

$$Po = 1/ [Nn\,(n - 1)]\,[\Sigma_{i=1,n}\ \ \Sigma_{j=1,k}\, n_{ij}^2 - Nn]$$

where,

N is the number of usability scenarios (cases)
n is the number of experts, which is 10 in this case
i represents each usability scenario that was used in the study
j represents the number of categories from which the users selected for the usability scenarios

$$Pc = \Sigma_{j=1,k}\, P_j^2$$

where,

$$P_j = 1 / Nn\,[(\Sigma_{i=1,n}\, n_{ij})]$$

Since kappa is normally distributed, the Z value was calculated using the formula below to compare the kappa values of this study with that of Andre's (1999) study.

$$Z_{obs} \quad = \quad (k_1 - k_2) / (\text{SQRT}(\sigma_1^2 + \sigma_2^2))$$

where, K_2 is the kappa obtained by Andre (1999), σ_1 is the standard error associated with K_1 and σ_2 is the standard error associated with K_2. The standard deviation for this study, σ_1, is calculated as the square root of the variance, which is calculated by the formula:

$$\text{Var}(k) = 2 / [Nn(n-1)] \frac{[\Sigma P_j^2 - (2n-3)(\Sigma P_j^2)^2 + 2(n-2)\Sigma P_j^3]}{(1 - \Sigma P_j^2)^2}$$

Scoring Expert Agreement

The participants' path through the User Action Framework was considered the primary data in this reliability study. The participants were asked to write down the classification path in the space provided between scenarios in the instruction sheet (Appendix G). The classification path taken by the participants was recorded and their final selection of the terminal node was documented for each of the fifteen usability problems. Consider the example shown in the Figure 4-3, where the sub-categories (the path) and the terminal node have been labeled. For this example, the path was translation > task structure and interaction control > preferences and efficiency > alternate ways to do tasks/specific short cut wanted. The first four levels were the sub-categories and the last level is the terminal node, which augments the classification with a description that matches the usability problem in the scenario.

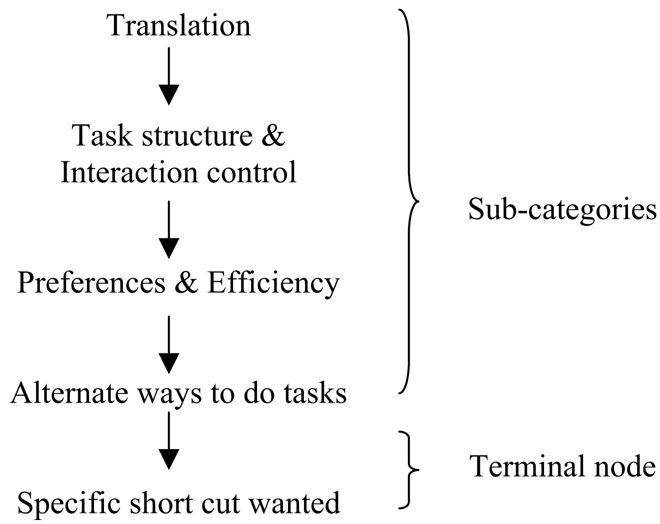


Figure 4-3: Path for a usability problem involving a translation issue

Because the UAF is comprised of a number of sub-categories at various levels (as many as 6 levels), agreement was calculated at each of the different levels within the hierarchical structure as well as overall agreement at all terminal nodes. For each usability scenario, the participant using the UAF is presented with a range of choices that are dependent upon the path taken to describe the problem. At the top levels of the UAF, the numbers of choices are usually small; typically the choices are between five or six items. The UAF broadens at deeper levels, presenting the user with as many as twelve choices at the lowest classification nodes. Therefore, the small differences in choices made early on result in large differences in terms of the number and kinds of choices faced later. As a result, the hierarchical structure of the UAF essentially holds up a higher standard for reliability because once two classifiers disagree, there is little or no chance for them to later reconverge to agreement.

RESULTS

Reliability measures, like kappa, are primarily used to calculate agreement among many users across a fixed number of categories. User agreement in the User Action Framework was analyzed in the following two ways:

1. Reliability at each level within the hierarchical structure
2. Overall reliability for the terminal node

Agreement at Different Levels in the UAF

Table 4-2 shows an example of the data from one usability case description. Scenario 8 was about the lack of a specific short cut for a particular task. Level 1 shows that all 10 participants agreed that this usability case description was a Translation issue because the scenario was about not helping the user how to do a particular task. At the next level (Level 2), 7 of the 10 participants agreed the scenario was about the Task Structure and Interaction control of the task. Three participants (#2, #3 and #5) thought the scenario involved Existence. In order to continue measuring agreement accurately, the data from participants #2, #3 and #5 were eliminated (indicated by strikethrough effect in Table 4-2) from further reliability measures since these participants were now taking a different path than the remaining seven. At Level 3, all 7 of the remaining 7 participants agreed that the issue was about Preferences and Efficiency. There were no more dis-agreements at other levels and all the remaining 7 participants agreed that the terminal node is about a specific short cut needed by the user. The example illustrated in Table 4-2 shows the approach for calculating reliability at different levels by eliminating participants that proceeded down a different path from the majority. This approach made sure that there were no continuous penalties for disagreement at lower levels when a participant was on a different path and had no opportunity to see the same choices as the other participants.

Table 4-2: Example classification of Scenario#8

Participant	Level 1	Level 2	Level 3	Level 4	Level 5
1	Translation	Task structure and interaction control	Preferences & Efficiency	Alternative ways to do tasks	Specific short cut wanted
2	Existence	~~Preferences & Efficiency~~			
3	Existence	~~Existence of a way~~	~~Existence of a feature~~		
4	Translation	Task structure and interaction control	Preferences & Efficiency	Alternative ways to do tasks	Specific short cut wanted
5	Existence	~~Preferences & Efficiency~~			
6	Translation	Task structure and interaction control	Preferences & Efficiency	Alternative ways to do tasks	Specific short cut wanted
7	Translation	Task structure and interaction control	Preferences & Efficiency	Alternative ways to do tasks	Specific short cut wanted
8	Translation	Task structure and interaction control	Preferences & Efficiency	Alternative ways to do tasks	Specific short cut wanted
9	Translation	Task structure and interaction control	Preferences & Efficiency	Alternative ways to do tasks	Specific short cut wanted
10	Translation	Task structure and interaction control	Preferences & Efficiency	Alternative ways to do tasks	Specific short cut wanted

Results of reliability calculations across all scenarios for various levels are given in Table 4-3. Column 2 indicates the number of cases analyzed for each level within the UAF. Depending on the scenario, participants had to traverse a number of hierarchical levels before reaching the terminal node. For example, some scenarios required navigation down to only the third level in the UAF before terminal nodes were presented. As shown in Table 4-3, 6 cases required Level 5 classification while only 1 case required classification at 6^{th} and 7^{th} level. Values in the Po column indicate the proportion of observed agreement while values in the Pc column indicate the proportion of agreement expected by chance. Kappa accounts for the fact that the proportion of chance agreement decreases as the number of choices increase. As shown in Table 4-2, the proportion of chance agreement is higher at the top levels in the UAF than the lower levels because there are fewer choices at the top of the framework. Observed agreement requires substantially higher values to overcome chance agreement at the top levels of the UAF. The kappa values shown in Table 4-3 (*k* column) indicated strong agreement at all levels within the UAF, especially at the top levels of the framework. The Z column contains the observed values for the standard normal variate obtained by dividing kappa by its standard error. The high z values indicated that kappa scores were significantly greater than chance agreement ($p < 0.001$). For Levels 5, 6 and 7 kappa was not calculated since only few cases were relevant at this level, limiting the number of data points for consideration. With such few cases, the agreement score would likely not be valid since the approximate normality assumption for kappa would be violated.

Table 4-3: Results of user agreement at each level in the UAF

Level	**Scenarios at this level**	**Po**	**Pc**	***k***	**Z*** **
1	15	0.986	0.346	0.979	23.48
2	15	0.911	0.163	0.893	39.16
3	15	0.910	0.084	0.902	66.49
4	13	0.617	0.056	0.594	52.83
5	6	--	--	--	--
6	1	--	--	--	--
7	1	--	--	--	--

*** $p < 0.001$
Note: Dashes indicate too few data points to calculate user agreement

Overall Agreement

Overall agreement among users was calculated by examining the terminal node descriptions across all usability cases. The overall agreement provides reliability information for the various paths taken by each classifier. Kappa results for overall agreement (Table 4-4) showed strong reliability ($k = 0.610$, $p < .001$), indicating agreement is greater than what would be expected by chance. In calculating kappa across all cases, the UAF is essentially transformed from six hierarchical levels into a flat structure with more than 100 terminal nodes. Therefore, the probability of chance agreement was extremely small (*Pc* = 0.051), considering the number of possible terminal nodes available to the users.

Table 4-4: Results of overall user agreement

Number of Scenarios	**Po**	**Pc**	***k***	**Z*** **
0.630	0.630	0.051	0.610	61.86

*** $p < 0.001$

Hypothesis Testing

Since kappa is normally distributed, it was possible to compare the overall agreement of this study with that of Andre's (1999). The hypothesis testing was done with an alpha level of 0.05 as follows:

$$H_0: k_1 = k_2$$
$$H_1: k_1 \neq k_2$$

The resulting Z score was calculated according the formula mentioned earlier in this chapter, as follows:

$$Z = (0.583 - 0.610)/ \text{SQRT} (9.7415E\text{-}05 + 8.8811E\text{-}05)$$
$$= -2.02$$

Hence, H_0 was rejected. Thus, it is supported that the overall agreement is significantly greater than that of Andre's 1999 study, ($Z = -2.02$, $p < 0.05$)

DISCUSSION

The results from the Reliability Study showed that the overall agreement ($k = 0.610$, $p < 0.001$) of the User Action Framework is significantly greater than demonstrated by Andre (1999), $Z = -2.02$, $p < 0.05$. Kappa across various levels (1-4) within the UAF ($k = 0.979$ to $k = 0.594$) showed that the agreement would be significantly greater than that expected by chance. Participants were especially consistent in using the parts of the Interaction Cycle to begin their classification of each usability problem. Only one participant, on one usability problem, selected a different part of the Interaction Cycle during the classification process. Such a result supports the notion that a theory-based framework is critical to providing a reliable classification system that helps build a shared understanding of the different attributes of a usability problem.

Landis and Koch (1977) provide some useful benchmarks to categorize strength of agreement according to the kappa statistic. The strength varies from "Poor" to "Almost Perfect" as kappa values vary from 0-1. Table 4-5 looks at various kappa values obtained in this study and categorizes them according to the scale provided by Landis and Koch (1977). These divisions provide some useful benchmarks to categorize kappa values, and interpret what it means in terms of users agreement. According to this, levels 1-3 obtained almost perfect agreement scores, with level 5 and terminal node obtaining moderate and substantial agreement scores, respectively.

Table 4-5: Strength of Agreement for various kappa values

Level	Kappa Value	Strength of Agreement
1	0.979	Almost Perfect
2	0.893	Almost Perfect
3	0.902	Almost Perfect
4	0.594	Moderate
Terminal-Node	0.610	Substantial

Built on a structured knowledge base of usability theories and concepts, the User Action Framework is intended to provide a common framework for usability practitioners to perform various usability engineering support activities, including consistent classification of usability problems. Consistent classification of a usability problem, based on its underlying cause in terms of the users' interaction with any application, is very critical to produce high quality problem reports. This will, eventually, aid in more efficient use of resources in the documentation process. In current practices, there is no standard procedure to convert usability data to specific design elements. Often, this "conversion" is done by means of brainstorming sessions involving experts from various areas like usability engineering, information architecture, graphic design etc. They use ad-hoc procedures, specific to a usability problem, to arrive at a design solution. It is the intention that the User Action Framework will solve this problem by allowing the Usability Practitioners to use the UAF as a common platform to arrive at direct design solutions.

A proper use of usability support tools, by interaction development groups, is possible only if they are used consistently and predictably from practitioner to practitioner (Andre et al., 2000). The results clearly show that the improved tool has higher reliability than would be expected by chance. Also, user agreement at terminal node was significantly higher than the previous study conducted by Andre (1999), thereby providing more support to use the UAF as a consistent framework.

CHAPTER 5. CONCLUSION AND DISCUSSION

The research set out to accomplish two main goals. First, to conduct a formative usability evaluation of the UAF Explorer, a component of the User Action Framework and its content. Second, to determine whether the User Action Framework still has good reliability since all the changes made after the previous work by Andre (1999).

The first part of this work focused on the formative evaluation of the Explorer interface. Two iterations of testing were conducted and usability problems were uncovered based on the observation of user performance while classifying some usability scenarios. Re-design recommendations were developed, which led a re-design effort to produce an improved UAF Explorer interface. The focus then shifted towards conducting the reliability study; participants classified a set of usability scenarios, and agreement among users was examined. The results showed that the User Action Framework demonstrated significantly better than chance agreement in all levels of the UAF. And, the agreement at the terminal node was shown to be significantly greater ($Z = -2.02$, $p < 0.05$) than that of the study by Andre (1999).

Interpretation of Results

Generally, the users, who participated in the formative evaluation studies, agreed that a tool like the User Action Framework would be very useful for usability practitioners and interaction developers to conduct a wide variety of usability engineering activities. Such activities, as perceived by the participants, are conducting usability assessments, classifying usability data, efficient project management techniques and also the possibility of using the User Action Framework to develop a methodology for consistent transformation of usability data into global and specific design elements. It was the objective of the developers of the UAF to build a suite of tools, encouraging the usability community to use the tool set as a common platform for interaction development and evaluation activities. The underlying theories and concepts, upon which the User Action Framework has been developed, would exactly support these kinds of activities. Hence, based on users' comments and perceptions, it can be concluded that the intended usage of the User Action Framework matched well with the users' mental model of tool usage. However, one participant did not quite agree with this concept, arguing that assessment and reporting activities should be specific to a particular application rather than using a universal framework for all applications.

A proper use of usability support tools, by interaction development groups, is possible only if they are used consistently and predictably from practitioner to practitioner (Andre et al., 2000). Without proper reliability, the result obtained by using the tool will depend on the evaluator and the data for the usability project will depend on the individual using the tool. An example of this variation can be seen in the heuristic evaluation technique presented by Neilson and Molich (1990). Heuristics did not provide a classification to differentiate between various usability problems (Dutt, Johnson & Johnson, 1994). This can be attributed to the structure of heuristics in that it is very general for detailed analysis with each heuristic covering a broad range of usability factors (Sears, 1997). This leads to overlaps among categories and also large gaps between categories, making it difficult to obtain reliable results. It has been noted (Jeffries, Miller, Wharton & Uyeda, 1991) that heuristics provide results that are not very successful in separating one problem description from another, leading to mis-classification and to problem identification that is not distinct. The User Action Framework was developed to solve this problem; the UAF is an essentially hierarchical structure of usability attributes, and unlike heuristics, with a description of specific usability problem description at the terminal nodes. Users traverse the UAF as a decision structure, selecting the most appropriate sub-category at each level of the hierarchy, and finally selecting a terminal node describing the usability scenario in question. An important goal for the UAF was to design in a model and structure of usability concepts and issues that usability professionals could use in a consistent manner. Results from the reliability study showed that users were in strong agreement when classifying a set of fifteen usability scenarios using the UAF. The high agreement scores at all levels in the

UAF showed the usability professionals can, indeed, use the User Action Framework for consistent classification of usability problems, based on the problem description at the terminal nodes augmenting the hierarchical structure of usability attributes

Limitations

One of the major limitations of any formative usability evaluation is that testing in a usability laboratory does not necessarily reflect the actual work context (Rubin, 1994; Whiteside et al., 1988). Though every effort was made to select a representative sample of users, there is a possibility that the participants that took part in the study may not also be a representative subset of actual users of the User Action Framework. According to Hollaren (1991), evaluators can also overestimate the power and generalizability of usability tests based on a small sample of subjects.

Practitioners would not deny the importance of providing consistent results from usability engineering support tools, but the issue of defining consistent performance is a different issue. Some practitioners may be interested in knowing that one evaluator can use a tool consistently across projects. Others may be more interested in knowing that different evaluators are relatively consistent in their use of the usability engineering support tool. Consistent use of the User Action Framework guarantees reliable classification of usability scenarios, but does not guarantee consistent quality problem reports that communicate problems and causes precisely. Reliability scores do not necessarily indicate consistent problem reports; it depends on both the structure of the usability framework that guides the description process and the content of the framework that helps to provide a complete understanding of the usability problem.

Recommendations for Future Research

Sophisticated Navigation techniques

As mentioned before, users in the formative evaluation studies identified the Windows Explorer paradigm as a rapid way to navigate the UAF content. However, they still had to click on the pluses and minuses several times before they could reach a specific sub-category. This proved to be time consuming and somewhat frustrating, especially to expert users. Future work can be directed towards developing sophisticated navigation methods using advanced information visualization techniques. Researchers have conducted research on various information visualization methods and have explained the importance of effective management of information spaces (Catarci & Cruz, 1996; Robertson et al., 1991)

Visually Chunking Information

Observations made during the study indicated that the website for the UAF is perceived as a text-heavy one. Research may be conducted to determine if chunking of information visually (i.e. by using different colors, tables etc) will offset the inherent text-heavy nature of the User Action Framework. A study by Venda and Beltracchi (1993) found that the chunking of data into an information model did improve the efficiency of the decision making process.

Number of Sub-category Choices

Some participants expressed concerns about the high number of sub-categories (up to 15 choices) to choose from, though it does not seem that the number of sub-categories affected the classification of usability problems. It might be worthwhile to look at ways to integrate two or more choices to produce fewer numbers of sub-categories to choose from at top levels and increasing the number of choices as users go down the hierarchical tree.

Content of the UAF

It will be fruitful to address the below mentioned areas in terms of improving the content of the User Action Framework. Examples could be added to all the sub-categories to help users map the usability scenario with a particular node. Research could be directed towards changing terminology of sub-category choices and refining cross-references to guide users to the desired part of the User Action Framework, by using appropriate “funneling techniques”. Finally, future work could address the issue of using technical terms (e.g. Cognitive Affordance) and simple phrases to balance both expert and novice users.

The work produced a useful and usable UAF Explorer interface and demonstrated that the User Action Framework can be used as a reliable platform to various usability engineering activities like classifying usability problems, generating usability reports, and converting usability data into global and specific design elements. This research also suggests that the User Action Framework can be used as a shared model for various usability attributes and as a better alternative for heuristic principles.

References

Andre, T. S., Belz, S. M., McCreary, F. A., & Hartson, H. R. (in press). Testing a framework for reliable classification of usability problems. In Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society.

Bennett, J.L. (1984). Managing to meet usability requirements, establishing and meeting software development goals. In Visual display terminals, usability issues and health concerns. (pp 161-184). Bennett, J., D. Case, J. Sandalin, & M.Smith. Eds. Prentice Hall

Catarci, T., Cruz, I.F. (1996). Information visualization, ACM SIGMOD Record, v.25 n.4, pp 14-15

Carroll, J. M., Kellogg, W. A., & Rosson, M. B. (1991). The task-artifact cycle. In J. M. Carroll (Ed.), Designing interaction: Psychology at the human-computer interface (pp. 74-102). Cambridge, UK: Cambridge University Press.

Carroll, J.M. & M.B. Rosson. (1985). Usability specifications as a tool in iterative development. Chapter 36 in Advances in Human – Computer interaction, Vol.1, Hartson, H.R. Ed., Ablex Publishing, pp 1-28

Carroll, J.M., Singley, M.K., & Rosson, M.B. (1992). Integrating Theory Development with Design Evaluation. Journal of Behavior and Information Technology.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46.

Cuomo, D. L. (1994). A method for assessing the usability of graphical, direct-manipulation style interfaces. International Journal of Human-Computer Interaction, 6(3), 275-297.

del Galdo, E., Williges R.C, Williges B.C, Wixon, D. (1986). An evaluation of critical incidents for software documentation design. In Proceedings of the Human Factors Society, 30th Annual meeting. Vol.1, pp 19-23

Doubleday, A., Ryan, M., Springett, M., & Sutcliffe, A. (1997). A comparison of usability techniques for evaluating design. In Designing Interactive Systems (DIS '97) Conference Proceedings (pp. 101-110). New York: ACM Press.

Dutt, A., Johnson, H., & Johnson, P. (1994). Evaluating evaluation methods. In G. Cockton, S. W. Draper, & G. R. S. Weir (Eds.), People and computers IX (pp. 109-121). Cambridge, UK: Cambridge University Press.

Flanagan, J.C. (1954). The Critical Incident Technique. Psychological Bulletin (July), 51 (4), 327 - 358

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378-382.

Garzotto, F., Matera, M. & Paolini, P. (1998). Model-based heuristic evaluation of hyper media usability. In Proceedings of the working Conference on Advanced visual Interfaces. AVI *'98,* pp. 135-145. New York: ACM Press.

Good, M., T.M. Spine, J.Whiteside, & P.George. (1986). Used derived impact analysis as a tool for usability engineering. In Proceedings of CHI ' 86, Human Factors and Computing Systems, ACM, pp 241-246

Hartson, H. R., Andre, T. S., Williges, R. W., & Van Rens, L. (1999). The user action framework: A theory-based foundation for inspection and classification of usability problems. In H. Bullinger & J. Ziegler (Eds.), Human-computer interaction: Ergonomics and user interfaces (Vol. 1, pp. 1058-1062). Mahway, NJ: Lawrence Erlbaum.

Hartson, H.R., & Castillo, J.C. (1998). Remote evaluation for post-deployment usability improvement. Proceedings of AVI '98 (Advanced Visual Interfaces). L'Aquila, Italy, 22-29

Hinderer, D., Kanter, L. (1998). Assessing Web Page Usability: Combining Techniques For Documentation And Product Usability Testing. In SIGDOC proceedings

Hix, D., & Hartson, H. R. (1993). Developing user interfaces: Ensuring usability through product and process. New York: John Wiley & Sons.

Holleran, P. A. (1991). A methodological note on pitfalls in usability testing. Behaviour and Information Technology, 10(5), 345-357.

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. M. (1991). User interface evaluation in the real world: A comparison of four techniques. In CHI '91 Conference Proceedings (pp. 119-124). New York: ACM Press.

Kanter, L., Rosenbaum, M., (1997) Usability Studies of WWW Sites: Heuristic Evaluation vs. Laboratory Testing. In SIGDOC proceedings. pp 153 –160. Salt lake City, UT

Kaur, K., Maiden, N.K. & Sutcliffe, A. (1999). Interacting with virtual environments: an evaluation of a model of interaction. Interacting with Computers, 11, 403-426.

Keenan, S. L. (1996). Product usability and process improvement based on usability problem classification. Doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Keenan, S. L., Hartson, H. R., Kafura, D. G., & Schulman, R. S. (1999). The Usability Problem Taxonomy: A framework for classification and analysis. Empirical Software Engineering, 4(1), 71-104.

Kirakowski, J., Cierlik, B. (1998). Measuring the Usability Websites. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting. pp 424 – 428. Santa Monica, CA: Human Factors and Ergonomics Society.

Landis, J.R., Koch G.G. (1977), "The measurement of observer agreement for categorical data," Biometrics, 33, 159-174.

Lee, A.T., Web site usability, usefulness, and visit frequency. (2000). In Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society.

Lim, K. H., Benbasat, I. & Todd, P. (1996). An experimental investigation of the interactive elects of interface style, instructions, and task familiarity on user performance. ACM Transactions on Computer- Human Interaction, 31-37.

Meister, D. (1985). Behavioral analysis and measurement methods. New York: John Wiley & Sons.

Nielsen, J., Landauer, T. (1993) "A mathematical model of the finding of usability problems," In *Proceedings of* ACM INTERCHI' 93 Amsterdam, pp. 206-213. The Netherlands, 24-29 April 1993

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In CHI '90 Conference Proceedings (pp. 249-256). New York: ACM Press.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), User centered system design: New perspectives on human-computer interaction (pp. 31-61). Hillsdale, NJ: Lawrence Erlbaum Associates.

Prescott, J., Crichton, M. (1999). Usability Testing: A Quick, Cheap, and Effective Method. In SIGUCCS proceedings. pp 176-179. Denver, CO

Rizzo, A., Marchigiani, E. & Andreadis, A. (1997). The AVANTI project: prototyping and evaluation with a cognitive walkthrough based on the Norman's model of action. In Designing Interactive Systems (DIS *'97*) Conference Proceedings, pp. 305-309. New York: ACM Press.

Robertson, G.G., Mackinlay, J.D., Card S.K. (1991). Cone Trees: Animated 3D Visualizations of Hierarchical Information Information Visualization. In Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems pp 189-194

Rubin, J. (1994). Handbook of Usability Testing. New York: John Wiley & Sons.

Sears, A. (1997). Heuristic walkthroughs: Finding the problems without the noise. International Journal of Human-Computer Interaction, 9(3), 213-234

Thompson, J. A. (1999). Investigating the Effectiveness of Applying the Critical Incident Technique to Remote Usability Evaluation. Unpublished thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

van Rens, L. S. (1997). Usability problem classifier. Master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Venda, V., Beltracchi, L. (1993) Data Chunking and Interface Display Design. In Proceedings of the Fifth International Conference on Human-Computer Interaction vol. 1 pp 648-653

Whiteside, J., J. Bennett, & K. Holtzblatt. (1988). Usability Engineering: Our experience and evolution. Chapter 36 in Handbook of Human-Computer Interaction. M.Helander, Ed., Elsevier Science Publishing company, Inc., pp 791-817

APPENDIX A. Application for IRB approval for Formative Evaluation

Virginia Polytechnic Institute and State University
Department of Industrial and Systems Engineering

Application for IRB approval for investigative project

Title of Project:	User Action Framework Formative Evaluation study
Principal Investigators:	Mr. Sriram Sridharan Dr. Tonya L. Smith-Jackson, Assistant Professor Dr. H. Rex Hartson, Professor

The Purpose of this Research

This research involves the study of the User Action Framework (UAF). The UAF is a methodology for organizing usability concepts and issues. This study involves experimentation for the purpose of evaluating and improving the UAF.

Procedures

The participants will be asked to perform a set of tasks using the UAF website. These consist of doing some tasks in the UAF website and classifying a set of usability problems using the UAF. All information collected will remain anonymous. Participants' actions will be noted and they will be asked to describe verbally their classification process. They also may be asked questions during and after the evaluation, in order to clarify our understanding of their evaluation. The session will last about 1 hour. The tasks are not very tiring, but the participants will be welcome to take rest breaks as needed.

Risks

There are no risks associated with this study other than those encountered from using a computer and a web-browser in everyday activities.

Benefits of this Project

Their participation in this project will provide information that may be used to improve the UAF. No promise or guarantee of benefits will be made to encourage them to participate in this research. If the participants would like to receive a synopsis or summary of this receive, they will be requested to contact the principal investigator.

Extent of Anonymity and Confidentially

The results of this study will be kept strictly confidential. Participants' written consent will be required for the researchers to release any data to anyone other than personnel working on the project. Any information collected will have their name removed and only a subject number will identify them during analyses and any written reports of the research.

Compensation

The participants will be paid $10 / hour for their time spent in the lab.

Freedom to Withdraw

The participants are free to withdraw from this study at any time for any reason without penalty.

Participant's Responsibilities

The responsibilities of the participants are given as follows:

- To notify the experimenter at any time about a desire to discontinue participation.
- After completion of this study, he / she will not discuss any experiences with any other individual for a period of one month. This will ensure that everyone will begin the study with the same level of knowledge and expectations.

For any questions about this research or its conduct, please contact:

Mr. Sriram Sridharan ssriram@vt.edu, 961-9091
Investigator

Dr. Tonya L. Smith-Jackson smithjack@vt.edu, 231-4119
Faculty Advisor

Dr. H. Rex Hartson hartson@vt.edu, 231-4857
Committee member

APPENDIX B. Informed Consent form for Formative Evaluation study

Virginia Polytechnic Institute and State University
Department of Industrial and Systems Engineering

Informed Consent for Participant of Investigative Project

Title of Project:	User Action Framework Formative Evaluation study
Principal Investigators:	Mr. Sriram Sridharan Dr. Tonya L. Smith-Jackson, Assistant Professor Dr. H. Rex Hartson, Professor

The Purpose of this Research

You are invited to participate in a study of the User Action Framework (UAF). The UAF is a methodology for organizing usability concepts and issues. This study involves experimentation for the purpose of evaluating and improving the UAF.

Procedures

You will be asked to perform a set of tasks using the UAF website. These consist of doing some tasks in the UAF website and classifying a set of usability problems using the UAF. We are not evaluating you or your performance in any way; you are helping us to evaluate our system. All information that you help us attain will remain anonymous. Your actions will be noted and you will be asked to describe verbally your classification process. You may be asked questions during and after the evaluation, in order to clarify our understanding of your evaluation. The session will last about 1 hour. The tasks are not very tiring, but you are welcome to take rest breaks as needed.

Risks

There are no risks associated with this study other than those encountered from using a computer and a web-browser in everyday activities.

Benefits of this Project

Your participation in this project will provide information that may be used to improve the UAF. No promise or guarantee of benefits has been made to encourage you to participate. If you would like to receive a synopsis or summary of this research when it is completed, please notify Sriram Sridharan.

Extent of Anonymity and Confidentially

The results of this study will be kept strictly confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research.

Compensation

You will be paid $10 / hour for your time spent in the lab.

Freedom to Withdraw

You are free to withdraw from this study at any time for any reason without penalty.

Approval of Research

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Industrial and Systems Engineering.

Participant's Responsibilities

I voluntarily agree to participate in this study. I have the following responsibilities:

- To notify the experimenter at any time about a desire to discontinue participation.
- After completion of this study, I will not discuss my experiences with any other individual for a period of one month. This will ensure that everyone will begin the study with the same level of knowledge and expectations.

Participant's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty.

_________________________ _________________________

Signature Date

Should I have any questions about this research or its conduct, I may contact:

Mr. Sriram Sridharan ssriram@vt.edu, 961-9091
Investigator

Dr. Tonya L. Smith-Jackson smithjack@vt.edu, 231-4119
Faculty Advisor

Dr. H. Rex Hartson hartson@vt.edu, 231-4857
Committee member

In addition, if you have detailed questions regarding your rights as a participant in University research, you may contact the following individual:

David Moore
Chair, University Institutional Review Board for Research Involving Human Subjects
Virginia Tech
Blacksburg, VA 24060
Ph: 540-231-5281

APPENDIX C. Pre-test questionnaire for Formative Evaluation study

Pre-test questionnaire for Formative Evaluation study

Directions: Please circle the relevant item or fill in appropriate response

Age: 18-28 29-38 39-48 49-59 60 and above

Gender: Male Female

Education (Highest level completed or attending)

2-Year College 4-Year College Masters Doctorate

Major field of study

Computer Science Human Factors Psychology Other (specify): ________________

For how long have you used a computer regularly? __________________________

Which computer platform do you use?

IBM-compatible PC Apple/Macintosh Other (specify): ________________

Which operating system do you use?

Windows 95/98 Windows 2000 Mac OS Other (specify): ________________

For how long have you used the Internet/Web regularly? __________________________

Which web browser do you use regularly?

Internet Explorer Netscape Navigator Other (specify): ________________

Rate your proficiency in using the Web?

Novice 2 3 4 5 6 7 8 9 Expert

Have you got any experience in usability engineering activities?

Yes No

If yes, rate your experience level

1-2 years 2-3 years 3-4 years 4 years or more

APPENDIX D. Participant instructions for Formative Evaluation study

Formative Evaluation on User Action Framework

Directions: You will be performing a series of tasks, which are given in the form of scenarios. The tasks should be performed in the order listed. You are encouraged to read aloud each task description before beginning the task. Also, please verbalize your thoughts by thinking aloud while performing the tasks. If the tasks are not clear or you are not sure how to perform them, please ask the test monitor.

Task 1. You are a usability practitioner in an industry. You recently heard about this online tool called user action framework and you get curious about what it does. So you decide to check it out by visiting the user action framework's website and exploring the interface.

Task 2. After you finish exploring the interface, you want to try classifying a usability problem. But you are not sure where your scenario fits in the various categories given in the website. So you want to search the website to see if you can find what you are looking for.

Task 3. You have just finished a usability evaluation and have a list of usability scenarios you want to classify. You know there is something called user action framework, which can be used to classify these scenarios but have not used it before. This time you decide to use this tool and hence visit the website to try it.

For each scenario, please write down your classification in the space provided below each scenario.

> **Scenario #1:** A user of a personal document retrieval system has been deleting numbered documents. The user now wants to reuse the old document numbers, but the system does not allow this.
>
> **Your classification**

Scenario #2: A user thinks he knows what he is doing on a certain task, but when he selects an object and clicks on an icon, he gets an error message. The problem is the error message is in a very small font and the color is too close to the background color, so he has difficulty reading the message. This problem is not about getting the error, but about the message received after the error

Your classification

Scenario #3: A database user accidentally deleted a number of related records. She knows that she can back out of this and correct the error, but the system doesn't help her find a way to do it. There is a button, labeled 'Back' for recovery from deletion but she was looking for something like 'Undelete' and didn't make the connection.

Your classification

Scenario #4: User knows generally that the Master Document feature of Word is used to allow treating several chapters in different files as a single document (e.g., for global editing). She wants to use this for her multiple thesis chapters, but the system doesn't help her figure out what she can do with it or how it might help her with her task. She has not yet done anything with it.

Your classification

APPENDIX E. Critical Incidents and User Comments in Formative Evaluation

Critical Incidents for 1st Iteration

1. User clicked on a cross-reference instead of clicking on a sub-category choice, subsequently getting the classification wrong
2. Confused by the word "Object" in various places
3. Users did not understand the difference between the words – "Tool Button" and "Tabs" in the home page
4. Clicked on "Upto Parent" and got thrown off
5. Error Messages Vs. Information Display under Assessment. User did not understand the difference, and hence got the classification wrong
6. Did not recognize a link
7. Clicked on "Information Display" instead of "Error Display"
8. Clicked on "Clarity, Precision…" instead of " Completeness, Sufficiency…"
9. Clicked on the navigation bar to start a new classification
10. Clicked on correct sub-categories (e.g., clicked on "Physical Actions")
11. Used the windows explorer paradigm to rapidly navigate the UAF web site
12. Noticed the navigation bar at the top of the page
13. User expects the links to be underlined in the web pages
14. Used an example given under a sub-category to identify correct classification path

User Comments for 1st Iteration

1. I do not know where I am
2. What I am supposed to do now? Have I reached the end?

3. I am not able to see the words in the home page inside the website
4. There is no difference between these categories in Content, Meaning
5. Oh, these are links! But they are not underlined
6. Say, something, like “You are Here”
7. The web site does not have any visual clue
8. There are lot of sub-categories under this
9. This tool can be used as a common framework to conduct various usability engineering activities
10. Classification should be based on problems for a particular application
11. The plus and minus are useful to navigate this web site
12. The web site seems to text heavy
13. I need to click on the plus and minus each time
14. More example in each sub-category will be very helpful
15. Typos reduces the credibility of a web site
16. The home page said deflector, but where is it?
17. I arrived at this by a process of elimination
18. I like the tree structure. It gives a very good overview
19. Looks like four tools are there. But I do not the see the names
20. Too much information on this web site. You have to do something about that
21. Have I finished classification now?

Critical Incidents for 2nd Iteration

1. User was not able to continue because he did not see the “Continue” button
2. Used the back button various times to start a new classification

3. User were able to distinguish between sub-category choices and cross-references

4. Identified the navigation bar because of its visual appearance

5. Users were able to identify their positions in the User Action Framework's hierarchical structure

6. Selected the correct sub-category – "Issues about Feedback"

7. Participants used the home page to identify the content inside the tool

8. Mapping of words in the UAF with particular scenarios, to aid classification

9. Users were able to recognize the links in the web site, because they were underlined

10. Clicked on "Upto Parent" to go back one level in the hierarchy

11. Used the plus and minus to have a look at the content

12. Used the Search capability to locate a sub-category in the UAF

User Comments for 2nd Iteration

1. How do I proceed from here?
2. There has to be an efficient way to do this
3. I like differentiating sub-categories and cross-references
4. There should be more examples in the sub-category choices
5. How to continue from this page?
6. "You are here" is very useful; this way I don't get lost in the web site
7. The use of colors to show the navigation bar is useful
8. Home page was useful to get an idea about what will be inside
9. Now, how to start another classification
10. There is a lot of information in the web site
11. I like the pluses and minuses in this frame; it is kind of a good idea to have it
12. I used the explanation to select this sub-category

APPENDIX F. Application for IRB approval for Reliability study

Virginia Polytechnic Institute and State University
Department of Industrial and Systems Engineering

Application for IRB approval for investigative project

Title of Project:	User Action Framework Reliability study
Principal Investigators:	Mr. Sriram Sridharan Dr. Tonya L. Smith-Jackson, Assistant Professor Dr. H. Rex Hartson, Professor

The Purpose of this Research

This research involves the study of the User Action Framework (UAF). The UAF is a methodology for organizing usability concepts and issues. This study involves experimentation for the purpose of evaluating and improving the UAF.

Procedures

The participants will be asked to perform a set of tasks using the UAF website. These consist of classifying a set of usability problems using the UAF. All information collected will remain anonymous. Participants' actions will be noted and they will be asked to describe verbally their classification process. They also may be asked questions during and after the evaluation, in order to clarify our understanding of their evaluation. The session will last about 1 hour. The tasks are not very tiring, but the participants will be welcome to take rest breaks as needed.

Risks

There are no risks associated with this study other than those encountered from using a computer and a web-browser in everyday activities.

Benefits of this Project

Their participation in this project will provide information that may be used to improve the UAF. No promise or guarantee of benefits will be made to encourage them to participate in this research. If the participants would like to receive a synopsis or summary of this receive, they will be requested to contact the principal investigator.

Extent of Anonymity and Confidentially

The results of this study will be kept strictly confidential. Participants' written consent will be required for the researchers to release any data to anyone other than personnel working on the project. Any information collected will have their name removed and only a subject number will identify them during analyses and any written reports of the research.

Compensation

The participants will be paid $10 / hour for their time spent in the lab.

Freedom to Withdraw

The participants are free to withdraw from this study at any time for any reason without penalty.

Participant's Responsibilities

The responsibilities of the participants are given as follows:

- To notify the experimenter at any time about a desire to discontinue participation.
- After completion of this study, he / she will not discuss any experiences with any other individual for a period of one month. This will ensure that everyone will begin the study with the same level of knowledge and expectations.

For any questions about this research or its conduct, please contact:

Mr. Sriram Sridharan — ssriram@vt.edu, 961-9091
Investigator

Dr. Tonya L. Smith-Jackson — smithjack@vt.edu, 231-4119
Faculty Advisor

Dr. H. Rex Hartson — hartson@vt.edu, 231-4857
Committee member

APPENDIX G. Participant instructions for Reliability study

Reliability Study on User Action Framework

Directions: You are given fifteen scenarios and each scenario represents a usability problem described in detail. Please use the UAF Explorer and classify each of these scenarios according to the structure of the User Action Framework. For each scenario, please write down the classification path in the space provided below each scenario.

An example of how to write down the classification path is provided as follows:

Important: Please do not refer back to the training materials as you complete the exercise.

<u>Example</u>

Scenario:

User clicks on a button to get the system to carry out a function and a confirmation message appears, "Are you sure you want to xyz?" As this was the only logical operation to the user at this point in the task, the user complained that the confirmation message was unnecessary and irritating and the system had forced the user to make an extra mouse click to deal with it.

Classification path:

1. A user thinks he knows what he is doing on a certain task, but when he selects an object and clicks on an icon, he gets an error message. The problem is the error message is in a very small font and the color is too close to the background color, so he has difficulty reading the message. This problem is not about getting the error, but about the message received after the error.

Classification path

2. User knows generally that the Master Document feature of Word is used to allow treating several chapters in different files as a single document (e.g., for global editing). She wants to use this for her multiple thesis chapters, but the system doesn't help her figure out what she can do with it or how it might help her with her task. She has not yet done anything with it.

Classification path

3. A user of a personal document retrieval system has been deleting numbered documents. The user now wants to reuse the old document numbers, but the system does not allow this.

Classification path

4. A drawing package has a very large number of functions, most of which are available to the user via a button bar crammed with various buttons. Each function is accessible via another way (e.g., a menu choice) as well, and our observations tell us that users mainly use the button bar for familiar and frequently used functions. So they don't usually have trouble figuring out which button to use; if it's not a familiar function, they don't use the buttons. This problem is about what happens when they do use the buttons. Because there are so many buttons, they are somewhat small and crowded together. This, combined with the fast action of experienced users, leads to clicking on the wrong button more often than users would like.

Classification path

5. A person using a Windows program for ftp file transfer wants to rename a file. She selects the file name and tries to type over it (as she does in the Explorer program on her PC), but this does not work. Eventually she figures out that you have to select the filename and then click the Rename command from a button bar. That leads to a small dialogue box with the filename in a text field where she can edit the name and click on OK. When she clicks OK, the system puts the new filename back into the list. She completes the task wondering why the system didn't provide a way to do the task directly.

Classification path

6. A user of a dbase-family database application had been deleting lots of records in a large database. The user knew that, in dbase applications, "deleted" records are really only marked for deletion (and can be undeleted) until a Pack operation is performed, permanently removing all records marked for deletion. At some point, the user did the Pack operation, but it didn't seem to work. After waiting what seemed like a long time (about 10 seconds), the user pushed the "Escape" key to get back control of the computer. As it turns out, the system *was* doing a Pack operation, which takes a long time for a large database. The user may have interrupted the operation with the "Escape" key, leaving things in an indeterminate state. If the system had let the user know it was, in fact, doing the requested Pack operation, he would have waited for it to complete.

Classification path

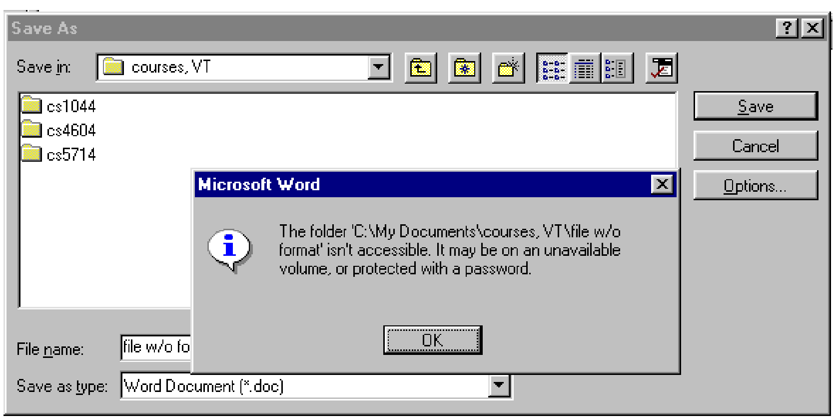
7. A database user accidentally deleted a number of related records. She knows that she can back out of this and correct the error, but the system doesn't help her find a way to do it. There is a button, labeled 'Back' for recovery from deletion but she was looking for something like 'Undelete' and didn't make the connection.

Classification path

8. A user of a home design program was ready to print his design and typed a control-P key combination, but it didn't work. After a pause to reflect, the user pulled down the File menu and saw the Print choice and, indeed, there was no "Ctrl+P" next to it. This problem is not about knowing what happened (or didn't), but it's about the system not allowing the user to perform the task the way she wanted.

Classification path

9. A user of Word has created a document that contains an outline of something. As the user saves it (with Save As), she names it "file w/o format" to distinguish it as the unformatted version. She was surprised by the resulting message: "Microsoft Word (with an "I" for "Information", not an "X" indicating an error) – The folder 'C:\My Documents\misc\file w/o format.doc' isn't accessible. It may be on an unavailable volume, or protected with a password." This message just seems so far off base that it's hard to make any sense out of it. It seems something is wrong and the file apparently hasn't been saved yet, but the system doesn't provide a way to tell what is wrong. (Aside: The real problem is that the file name used contains a "/" character, not allowed in Windows. These apparently made it look like part of a path name to Windows and it went off on the wrong track, confusing the user.)



Classification path

10. In an apparent attempt to be complete, an error message is long; it just carries on. The extra words and explanation obscures the simple message that the system is trying to convey to the user. The user ends up being confused and irritated.

Classification path

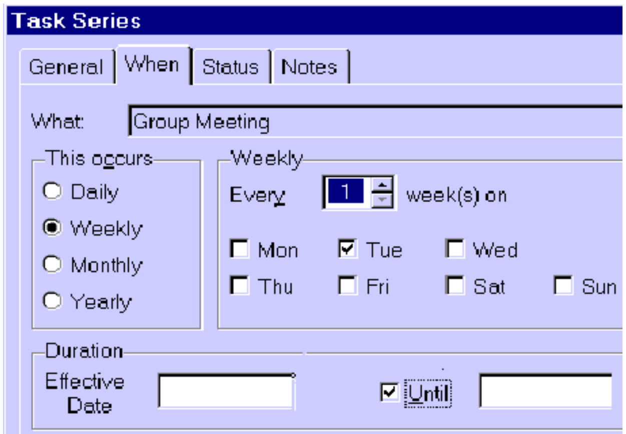
11. A studly dude had a stereo with a CD player that had controls (for Play, Stop, etc.) that were embossed black icons on a black background. This black-on-black motif is cool and sexy, but had an impact on usability. You could see the raised embossing of the icons in good light, but this dude liked music in a low-light ambiance, a condition that made it very difficult to see the icons. As an experienced user, he knew that he should push the Play button when he wanted to play a CD, but he couldn't always tell for sure where that button was on the plain black front of the CD player.

Classification path

12. In a word processing task, while trying to fit a small document just on one page, the user selected all the text and went to the font size menu. Seeing only 10 point and 12 point choices on the menu, she picked 10 but that made the document a bit too small on the page. She commented that it would have been nice if an 11-point font had been available and went on to other parts of the task. The evaluator/observer noted that, in fact, an 11 point font could have been selected, by typing '11' into the little text box part of the menu where the current font size selection is shown, but the system did nothing to help the user know about this possibility.

Classification path

13. User is filling out an on-line form and gets to a field for a date, but there is no indication about what format to use. The user tries something and gets an error message and then is able to correct it and get the system to recognize the date. Some affordance is given through the label that says it is a date value, but it still doesn't allow her to get the format right the first time.



Classification path

14. In Word on the PC, when you use drag and drop to move text and have to go outside the text that shows on the screen, it scrolls when you get to the top or bottom. Unfortunately, the speed of scrolling is controlled only by the speed of the machine and ends up being too fast for the user to control. The result is thoroughly intimidating and frustrating. The system has put him in a difficult spot, having to hold the mouse button depressed, with the text attached to the cursor, going back and forth not able to find a place to put it.

Classification path

15. A vision-impaired user has difficulty reading the message text in a dialogue box. This is a more or less expected occurrence for a visually impaired user and is not the problem in itself. The problem here is that the system doesn't help the user find a way to set the font to be larger and bolder or to get an alternative audio version of the messages.

Classification path

APPENDIX H. Informed Consent Form for Reliability Study

Virginia Polytechnic Institute and State University
Department of Industrial and Systems Engineering

Informed Consent for Participant of Investigative Project

Title of Project:	User Action Framework Reliability Assessment
Principal Investigators:	Mr. Sriram Sridharan Dr. Tonya L. Smith-Jackson, Assistant Professor Dr. H. Rex Hartson, Professor

The Purpose of this Research

You are invited to participate in a study of the User Action Framework (UAF). The UAF is a methodology for organizing usability concepts and issues. This study involves experimentation for the purpose of evaluating and improving the UAF.

Procedures

You will be asked to perform a set of tasks using the UAF database. These tasks consist of classifying a set of usability problems using the UAF. We are not evaluating you or your performance in any way; you are helping us to evaluate our system. All information that you help us attain will remain anonymous. Your actions will be noted and you will be asked to describe verbally your classification process. You may be asked questions during and after the evaluation, in order to clarify our understanding of your evaluation. The session will last about 1 hour. The tasks are not very tiring, but you are welcome to take rest breaks as needed.

Risks

There are no risks associated with this study other than those encountered from using a computer and a web-browser in everyday activities.

Benefits of this Project

Your participation in this project will provide information that may be used to improve the UAF. No promise or guarantee of benefits has been made to encourage you to participate. If you would like to receive a synopsis or summary of this research when it is completed, please notify Sriram Sridharan.

Extent of Anonymity and Confidentially

The results of this study will be kept strictly confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research.

Compensation

You will be paid $10 / hour for your time spent in the lab.

Freedom to Withdraw

You are free to withdraw from this study at any time for any reason without penalty.

Approval of Research

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Industrial and Systems Engineering.

Participant's Responsibilities

I voluntarily agree to participate in this study. I have the following responsibilities:

- To notify the experimenter at any time about a desire to discontinue participation.
- After completion of this study, I will not discuss my experiences with any other individual for a period of one month. This will ensure that everyone will begin the study with the same level of knowledge and expectations.

Participant's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty.

_________________________ ____________________________

Signature Date.

Should I have any questions about this research or its conduct, I may contact:

Mr. Sriram Sridharan ssriram@vt.edu, 961-9091
Investigator

Dr. Tonya L. Smith-Jackson smithjack@vt.edu, 231-4119
Faculty Advisor

Dr. H. Rex Hartson hartson@vt.edu, 231-4857
Committee member

In addition, if you have detailed questions regarding your rights as a participant in University research, you may contact the following individual:

David Moore
Chair, University Institutional Review Board for Research Involving Human Subjects
Virginia Tech
Blacksburg, VA 24061
Ph: 540-231-5281

VITA

SRIRAM SRIDHARAN

Relevant Skills

Operating systems	Windows 98/00Me/NT, Mac OS, UNIX
UI tools & packages	PhotoShop, Flash, Dreamweaver, Fireworks, VB, Visio & more
HCI techniques	Think Aloud, Design Walkthrough, Design Prototyping & more
Human factors skills	Task analysis, Needs analysis, Mental modeling & more
Knowledge of	C, C++, HTML, DHTML, Director, JavaScript, SAS & more

Career Profile

Strategy & Design of User Interaction – backed by Usability Engg Methodology & Guidelines

Usability Engineer / **User Interface Designer**, Virginia Tech **Dec '99 – present**
Evaluating and re-designing a website for a suite of usability engineering support tools
Techniques/Tools: Think Aloud, Paper Prototyping, Design Walkthrough, Dreamweaver

Lab Assistant, Faculty Development Institute **May '01- Aug '01**
Assisted faculty members design web-based course materials by applying principles of Usability, Visualization, Functionality and Accessibility
Tools taught: Dreamweaver, Adobe products, Flash, PowerPoint

Usability Intern, User Works, Inc **May '00 – Aug '00**
Series of usability testing for various projects and re-design efforts. Projects include:
- Co-discovery on a web-based expert application designed for "smart car-shoppers"
- Wireless web features of mobile phones involving real users through a focus group study
- Usability evaluation of Verizon InfoSpeed DSL Self-Installation process with real users

Techniques/Tools: Think Aloud, Mental Modeling, Heuristic principles, PowerPoint

Graduate Research Assistant, Assessment Lab **Sep' 99 – Dec '00**
Research on designing interfaces for disadvantaged users and universal usability
Responsibilities: Conducting experiments, collecting and analyzing data

Other Projects

- Designed and evaluated prototype for a graphical touch-screen interface for a copier. Identifying user classes, users/client needs, usability testing and subsequent re-design
- Evaluated the usability of touch-screens in Exxon gas pumps and recommended a more interactive design by applying heuristic evaluation principles

Education

Academic Training aimed at Practical Setting

M.S in **Human Factors Engineering**, Virginia Tech, **GPA: 3.52** (major field) **Dec 2001**
Courses: Usability Engineering, User Interface Design, Information Visualization, Research Design, Systems Design, Human Information Processing, Training Design, Macroergonomics
Thesis: Usability and Reliability of User Action Framework: A Theoretical Foundation for Usability Engineering Activities
Techniques/Tools: Think Aloud, Critical Incidents, Design Walkthrough, Dreamweaver

B.S in **Chemical Engineering**, University of Madras, **GPA: 3.70** **May 1999**

Professional Activities

- Student Volunteer, UPA Tenth Annual Conference, Lake Las Vegas, NV, June 25-29, 2001
- Attended the ACM Conference on Universal Usability, Arlington, VA, Nov 16-17, 2000

Attributes

- Proven research & communication skills, reliable as both individual & team player
- Self-motivated, innovative and attentive to detail
- Comprehensive training in human factors analysis, design and evaluation methodologies

Affiliations HFES, UPA, SIG-CHI - Bay CHI, GB CHI, ACT

For detailed resume, visit http://filebox.vt.edu/users/ssridhar/resume.html