

Michael Adcock, Khue Duong, Marisa Haberfelde, Ann Swearingen  
TC 518  
May 28, 2008

## **Usability Evaluation: “TiddlyWiki Thesaurus”**

### **Introduction**

For this investigation we recruited four participants, all second-year UW Information School graduate students in the MLIS program. Two of them are currently engaged in building a thesaurus for LIS 537, Constructing an Index Language, and two of them completed the project in previous quarters (spring and fall 2007, respectively). They all have or had the same instructor, course materials/readings, and project requirements. Two of the participants were already somewhat acquainted with a current version of the TWT. One of the other participants has a degree in computer science, and all four were very computer- and internet-savvy.

We held the usability evaluations over a two day period (May 25-May 26, 2008), in a lab in Mary Gates Hall. The sessions were run by two team members, who coordinated taking notes, managing the paper prototypes, and facilitating the interviews and conversations. Each session took about an hour to complete. Refreshments were offered, and gift cards were provided as incentives. Each participant was fully briefed at the beginning of the sessions as to what would take place, and was asked to sign a consent form. The sessions were audio-taped for future reference.

### **Procedure**

Each session was guided by a set of protocols laid out in an administrator script, to ensure consistency in the proceedings of each team.

The evaluation consisted of five segments:

1. After introductions were made, the participant was asked to work through a series of four simple tasks with each of the paper prototypes, using a think-aloud protocol.
2. After testing each prototype, they were asked to complete a short evaluation survey in which they ranked the features involved in each task on a seven-point Likert scale.
3. Next, they were asked to draw a picture of their ideal thesaurus building tool using pen and paper, with no time limits given.
4. In a final interview, they were asked to explain the features and functionalities of their sketch.
5. Finally, they were asked to verbally assess the prototypes and add any other comments, suggestions or questions.

### **Methods**

We chose to use low-fidelity paper prototyping for several reasons. It was the most practical evaluative tool in this context: it is portable, inexpensive, and relatively easy to reproduce and modify. Because it was a low-fidelity representation, it let us focus on essential features of the thesaurus application, in this case issues which emerged as priorities in the earlier investigation: navigation and terminology. This method highlights these problem areas, permitting both investigators and participants to see past any design “clutter” to questions of function and basic features. (As one participant noted, it let him “get inside” the tool).

Two prototypes were demonstrated, each consisting of four frames. Changes to the existing TWT design were made, to the navigation bars, terminology and layout. These included dropdown menus, simplified side

navigation, renamed controls, relocated search, save and login functions, and new “housekeeping” features. The second prototype also featured vertical panels which could be expanded and collapsed, allowing users to customize their views of what they are working on at a given time. Users have indicated a need to work with content in different views, and to see their work in juxtaposed panels.

The post-task surveys were based on the results of the recent heuristic evaluations, as well as the earlier cognitive walkthrough. The surveys tacitly encouraged the participants to focus on navigation, vocabulary and layout issues which had emerged in those earlier investigations. Asking them to do a survey encouraged them to think about those issues in relation to each prototype.

The sketching exercise was designed to elicit original critiques and suggestions from the participants. The participants were given pens and unlimited sheets of 8 ½” X 11” white paper, and given unlimited time to draw. They were encouraged to talk while they worked, if they wished to, and were allowed to annotate the drawings in their subsequent explanations of their features. Sketching allowed the participants to engage with the application in a more concrete way, and to productively expand on their critiques of the prototypes.

The final interview questions enabled participants to focus on the same issues raised in the earlier survey questions, but in relation to their own sketches.

## **Findings**

Several broad themes emerged in the sessions, which both confirmed and expanded on earlier discoveries. Areas of note were terminology, functionality, user process, and feature requests.

### **Terminology**

- Participants found some of the terms to be unclear or confusing. They consistently identified problems with the same terms, and made similar suggestions for modifications. “Error checking” and “build” were disliked by two of the users, and one user perceived “help” and “error checking” as having negative connotations.
- “Save locally” and “save to server” created confusion, though participants liked having both options. The term “download” was suggested as an alternative for “save locally.”
- “Track changes” was popular as a feature, but one participant suggested that it be renamed “edit history” and others weren’t clear on what it meant. The same individual also assumed it was analogous to the Microsoft Word “track changes” feature, and could be toggled on and off as an editing tool.

### **Functionality**

- All of the participants commented on the structure of the “help” dropdown menu. Two wanted the glossary function moved to the side navigation bar because it would be more convenient there.
- They wanted to see the alphabetical and classified schedules broken out of their dropdown menu and split out on the top navigation bar.
- Three of the users expressed a need to be able to work/edit in one window while being able to see changes to the classified schedule in an adjacent window.
- “Error checking” was a popular feature (despite objections to the term itself). Users expressed a desire to check their work after making changes or adding terms to the application.
- One participant pointed out the need for feedback and confirmation after performing tasks such as edit, save, add and delete.

- How to select terms for editing was unclear. Participants wanted to know how they would “get at “a term, if they could click directly on a hyperlinked term into an editing space, or if they would need to search the thesaurus to select and then edit a term.
- Participants generally liked the TWT.

### **User Process**

- Participants identified several functionality issues related to the steps involved in performing some basic tasks.
  - All of the participants were anxious about the lack of clarity about deleting terms. (“The delete button is scary!”)
  - The steps for selecting a term to delete were unclear, as were the expected result. (E.g., would it delete all of the terms, the thesaurus as a whole, a single term?)
  - The process of selecting a term to be deleted was also unclear, whether it could be selected directly from the schedule, or if the search function would need to be used.
  - Two participants asked about a mass-delete option.
- Concerns about saving their work compounded their worries about the deletion process. All the participants noted that they would save their work frequently, and saw saving it to the server as more reliable than saving locally.
- Participants wanted to be able to add terms in more than one way:
  - As single terms
  - As imported spreadsheets
  - Via a suggested “quick add” function.
- One user found the search function unclear, in that it was difficult to differentiate whether it would search only the thesaurus content, or the application as a whole

### **Feature Requests**

- Automatic backup/save functionality.
- Enhanced support for online collaboration, through an expanded “track changes” function which would let users follow their team members’ work in a distributed environment, such as online versions of the project. This would include the option to annotate terms with comments and source information, to sort by different elements, and to accept/modify changes made by other team members.
- A “quick add” feature, located in a fixed pane.
- “Sort by” functions within hierarchy display (for example: sort by date added, source, etc.).
- “Drag and drop” text from workspace to hierarchy list, or in chunks from one part of the hierarchy display to another, in order to represent hierarchical meaning relationships visually.
- Incorporation of colors into facet display, to support sorting of broad concepts and terms through a type of associative visual “coding”.
- Administrative options which would permit customization; a modifiable interface.
- A separate toolbar for all of the functions which modify terms (add, edit, delete, save).

### **Redesign Implications**

- Specific terminology related to critical tasks needs to be reworked.
- The tool needs to support online collaboration more effectively.
- The search function is ambiguous, and seems to be of little value when building a thesaurus.

- The user needs to be able to look at the work he is doing on the classified schedule from different perspectives
  - Zoom in on a specific term
  - Pull back and view the entire structure of the thesaurus
- People think of the relationships between terms visually. Meaning-relationships within the classified schedule need to be represented—and manipulated—visually.
- Graphical elements such as color coding, pictures and diagrams would help users identify groups of related terms.
- Users see the classified schedule as the focal point of the tool, and the main way to manage concepts.
- Care should be taken not to lose the current strengths of the design.