8. Software tool evaluation

8.1 Introduction

This chapter discusses the results obtained through an user evaluation study of the PhiloSurfical tool (presented in chapter 6).

The chapter is organized as follows: first we give an overview of the purpose of the experiment and describe the details of the experiment design; secondly, we present all the data gathered; finally, we proceed to the discussion of their significance, thus generating a detailed outline of the requirements for future versions of PhiloSurfical.

8.2 Purpose of the experiment

The development of the PhiloSurfical tool, in the general context of our thesis' work, mainly served as a way to test the ontology capabilities and the feasibility of its use within a real-world system. Consequently, the tool has not been fine-tuned to the extent need for a task-based evaluation.

For this reason, we carried out instead a user-interaction evaluation experiment, aimed at testing our prototype's usability. More precisely, we expected to gather results with respect to three main points:

1) *Generic ease of use of PhiloSurfical*: do people understand its scope and context? Do they manage to navigate around the various sections and perform the expected operations? Do they find it attractive and engaging?

2) *Perception of the underlying ontology*: do people realize it is an ontology-based application, and if yes, how does that matter to them? Are the most 'obscure' references to the ontological representations well hidden from the 'naive' users, or are they of impediment for a successful user-interaction process?

3) *Familiarity with the learning-narrative idea*: do people find it easy to understand? Do the interface provides a simple way to trigger the narrative mechanisms? Are the types of narratives which are available useful to the users?

8.3 Experiment design

In order to carry out the experiment we chose 5 people in our university department, making sure all of them had little or no philosophical competencies. We then asked them to accomplish some common tasks we thought were easily achievable by using our prototype tool. In order to gather information about their reactions, we asked them to think aloud and talk us through their moves and decisions.

The experiment was set up on a computer running a software which allowed us to store an audio/video recording of each evaluation session. At the end of all

the sessions, we analyzed the five videos and took down a detailed list of the comments the volunteers made. Once the video annotations had been completed, we studied the gathered data in order to identify important trends and problem categories.

As mentioned, all the volunteers were chosen because they were not familiar with the philosophical domain. By doing so, we wanted to avoid the situation where respondents would have been facilitated in completing one of the tasks (e.g., answering a question like "what is the relationship between rationalism and empiricism?") because of their pre-existing philosophical knowledge. Moreover, out of five volunteers, only two of them were working in a Semantic Web-related field. As a consequence, it is also fair to suppose that more than half of the subjects are not familiar with typical SW concepts and technologies, such as 'ontology' or 'ontology-based navigation'. We believe that this feature also helped in avoiding a situation where users are already 'experts' with respect to the technology tested.

8.3.1 The tasks

We gave the respondents five different tasks. In general, the tasks were conceived so to have volunteers use the various sections (tabs) of PhiloSurfical. In particular, task 1 involves the first tab (splash screen), task 2 the second tab (which is titled 'browse the text'), task 3 the third tab ('browse the annotations'), tasks 4 and 5 the fourth tab ('browse the pathways'). There is only one remaining tab, the fifth one ('browse the ontology'). However this does not provide any facility linked to the philosophical text's navigation; in fact, we

added it to the prototype mainly as a way for Semantic Web and knowledge representation experts to have a glance at the underlying ontology supporting the system. PhiloSurfical is intended to be a showcase of the *functionalities* enabled by the ontology - not an attempt to make people interact directly with it. In fact, future and non prototypical versions of the tool are not likely to include the 'browse the ontology' tab anymore. In conclusion, since this tab does not provide any 'learning' functionality, we decided not to include it in the evaluation. The tasks' descriptions given to the respondents are the following:

Task 1) Play with the software, think aloud and try to discover what is it all about in five minutes.

Task 2) Retrieve sentence 3.01 of the Tractatus, highlighting its different translations

Task 3) Find as much information as you can about the wittgenstenian concept of 'object', e.g., where it appears in the text, how it relates to the other ideas of Wittgenstein and especially how it differs (if it does) from the wittgenstenian concept of 'thing'.

Task 4) Gather information about the 'problem of the foundation of mathematics' (e.g., what it is, what are the philosophies tackling it, where can relevant web resources be found).

Task 5) What's the relationship between the 'philosophy of Frege' and the 'philosophy of Russell' ?

8.4 Experiment results

In table 8-1 we can see all the experiment results, organized according to the tasks' sequence.

The central column gathers respondents' verbatim comments and feedback; in the right column we are reporting the annotations we made during the volunteers' performances. When the same comment was made by more than one respondent, we indicate it in brackets after the comment's text.

Apart from task 2 (which has been unanimously judged as straightforward), all the other tasks had been successful in stimulating respondents' feedback. In the next section we will discuss such results in order to draw some specific requirements for the next version of PhiloSurfical.

TASK	Users' Comments	Our Comments
1: Have a play with the software, think out loud and try to discover what's it about in five minutes	• the splash screen is really useful in explaining what PhiloSurfical is and its purposes *(4 volunteers)* • the FAQ section is interesting • the layout is beautiful and clear • 'browse the ontology' is useful! • a search box is totally missing, usually I associate 'browsing' with 'searching' *(2 volunteers)* • the ontology tab is confusing - why is it there? • for a philosophy neophyte, the text can be quite obscure and mysterious - especially in the 'browse the text' tab, it's not clear whether it's all the text, or not ('browse the Tractatus' might be clearer) (*2 volunteers*) • add a link to a webpage showing examples of how to use the system • state more clearly that this is an example of PhiloSurfical with a real book, the Tractatus • the help icon on the first page should work! • give a description also of the difference among the various tabs' purposes and functionalities	• the help system has not been used by any of the volunteers!

TASK	Users' Comments	Our Comments
2: Retrieve sentence 3.01 of the Tractatus, in different translations	• it's very easy to find the sentence and its translations *(5 volunteers)*	• the task is straightforward for everybody
3: Find as much information as you can about the wittgenstenian concept of 'object', e.g., where it appears in the text, how it relates to the other ideas of Wittgenstein and especially how it differs (if it does) from the wittgenstenian concept of 'thing'.	• it is hard to understand what the colored buttons stand for (D/I/A); a tooltip or popup is needed *(4 volunteers)* • I went straight to the pathways-tab thinking that it was the right place to search for information, but then I got stuck • a search box is missing, therefore I try to use the browser search function *(2 volunteers)* • I would like to create more complex queries, e.g., for retrieving all paragraphs dealing with concept X AND Y (and so on..) • the system should say somewhere that if an item is not findable in the 'categories' you can still look for it in the ontology! • the help icons are easily confused with the titles, they shouldn't stay on top; also, they should have their different topics mentioned • in the 'inspect' panel, provide explanations about which are the possible relations among ideas, by using the help or maybe with a tooltip • need of a back/forward mechanism for all the tabs	• people try to use the ontology-tab because they misinterpret the 'concepts' mentioned in the task; they think they are classes in the ontology, instead of instances of the class 'concept' used for annotating the text (*2 volunteers*) • the interface is too complicated, users tend to get lost and miss the important buttons
4: Gather information about the 'problem of the foundation of mathematics' (e.g., what it is, what are the philosophies tackling it, where are other web resources about it).	• the pathways idea comes through quite clearly • pathway results look good and interesting • the 'recent-items' tab is very useful, I usually like to keep track of what I've done • I would like to have a search function on all the instances related to the text • the 'change' button is not immediately perspicuous, substitute it with something bigger and more explicit e.g., 'pick another item' (*3 volunteers*) • if I click on any item this is automatically put back 'into focus', but I could not notice this. Something must tell me what happens (*3 volunteers*) • I would like to have a pathway giving the generic description of something • the pathways names' are not clear, they are too specific. Rephrase them in a more obvious way for non-competent users.. • the title 'pathways list' should be changed into something clearer, e.g., 'search an item using the pathways' • the java applet is sometimes not working	• People seem to understand the functionalities of the tab, but there are various titles and naming issues to fix.

TASK	Users' Comments	Our Comments
5: What's the relationship between the 'philosophy of Frege' and the 'philosophy of Russell' ?	• I cannot associate the term 'view', used in the taxonomy, to meanings such as philosophy or theory (*3 volunteers*) • it seems that I could type into the 'item in focus' box, but it doesn't let me. Quite counter-intuitive! (*3 volunteers*) • I need a search box (*2 volunteers*) • the 'instance-info' link should become another pathway! (e.g., 'generic information about an instance') • on the 'change-item' popup: we need some help/tooltips here too. • on 'recent items and search' tab, there is no search! • I need some help about the categories in the 'change item' popup	• When searching for 'philosophy of Frege', people are looking for the 'person' instances first, instead of searching directly for instances of 'view'. (*3 volunteers*)

Table 8-1. Summary of the results of PhiloSurfical's user evaluation

8.5 Discussion and requirements' definition

In order to discuss the results in a more structured way, we will go through each of the three categories introduced above when outlining the purpose of the experiment. Therefore we have:

1) **Generic ease of use of PhiloSurfical**:

[Do people understand its scope and context? Do they manage to navigate around the various sections and perform some operations? Do they find it attractive and engaging?]

All the people claimed to get the idea behind PhiloSurfical in a straightforward manner. Also, some of them noticed how the information page is providing various links which are of fundamental importance for understanding the context

of the software prototype. The interface was seen to be appealing and easy to navigate. However, in various cases respondents indicated some points where the user interface was unclear or hard to use. We have summarized these remarks in the following points, which can also be seen as requirements for future versions of the tool:

1. need to add a generic 'search box' functionality

   • include search-options e.g., search for 'all instances related to the text', or more complex searched e.g., search for 'sentences about concept x AND y'.

   • specify the meaning of 'search' in the section 'Recent items and search'

2. improve the help system, in particular

   • add an 'example' page with screenshots, detailed how-to etc.

   • modify the help icons, so that they become flashier and more specific to the section they belong to (e.g., 'help about ...')

   • include a description of the different tabs' intended functionalities

   • add some help to the 'change item' popup in tab4

3. make sure all the tabs and panel titles are simple to understand, especially by testing them with non domain-experts

4. need of a back/forward mechanism, specific to the various panels

5. add more tooltips to tab3, so that it would become easier to understand the intended meaning of the different panes

6. improve the user interaction on tab4, especially when a user wants to explore new pathways by selecting a new item (this requirement is discussed in more details below, when discussing point 3)

2) **Perception of the underlying ontology**:

[do people realize it is an ontology-based application, and if yes, how does that matter to them? Are the most 'obscure' references to the ontological representations well hidden from the 'naive' users, or are they of impediment to their productivity?]

The majority of people did not pay too much attention to the fact the application runs on top of an ontology, but in general they mentioned it as an interesting feature.

One person claimed that tab5 (the 'navigate the ontology' tab) is very interesting and fundamentally helps in understanding the tools capabilities. However, 4 out of 5 respondents claimed that tab5 is confusing: they do not see clearly its functionality and treat it mistakably as another search mechanism, or as a way to query the knowledge base. It is important to remember that this negative reaction was not caused by the specific classes or relations specified in the ontology, that is, respondents never commented on the 'quality' or 'correctness' of the philosophical ontology. Rather, this confusion was a consequence of the purely explicative (i.e., non functional) role the ontology has in the context of the PhiloSurfical prototype (as discussed also in section 8.3.1) - in other words, users just did not understand what to do when presented with such a complex object.

Moreover, on tab3, they tended to overlook the categories used for organizing the various annotations about Wittgenstein; also, they did not perceive them as ontological classes. Sometimes they complained that certain category-names (e.g., *view* or *concept*) are counter-intuitive and do not summarize well what are the entities they stand for (e.g., philosophies, or simple ideas related to Wittgenstein).

To summarize these results, we outlined the following requirements:

1. Tab5 should be either removed or presented in such a way that its purely explicative purpose becomes immediately clear.

2. The classes called *view* and *concept*, at least on the PhiloSurfical interface, should be presented using a name which is less prone to misinterpretations. In the ontology, instead, they might keep the original name.

3. The categories presented in tab2 should be easily associated to the relevant classes in the ontology; also, by doing so, users should understand that the information in the PhiloSurfical knowledge base has a larger scope that what is covered by the categories in tab2.

3) **Familiarity with the learning-narrative idea**:

[do people find it easy to understand? Do the interface provides a simple way to trigger the narrative mechanisms? Are the types of narratives available useful to the users?]

In general, people approached the 'learning narratives' tab as a special search facility. After a few tests, they understood the features of a structured search, claiming also that the pathways' description was of fundamental help for understanding what was going on.

However, they also noticed various problems.

First of all, sometimes the pathways' descriptions (and also their names) are convoluted; as a result, they claimed that a person who is not a philosophy expert would have difficulties in getting their intended meaning.

Secondly, they noticed some problems related to the updating of the interface when new contents are selected, problems that caused the *iterative* pathways' selection process to become quite unintuitive.

Thirdly, they had some difficulties during the selection of new contents from the knowledge base, when trying to identify contents of interests so to start a new pathway-search.

Finally, for that regards the variety of pathways available, the most relevant comment is about the lack of a 'route' which takes directly from a person (e.g., a philosopher) to his/her intellectual products. Please notice that this topic emerged also during the ontology evaluation (cf. section 7.6.9). A related comment concerns the fact that 'instance-info' button (added as a way to see, at any time, the complete set of data available for an instance) should become instead a pathway, providing generic information about the selected instance.

So, to summarize, we have the following requirements:

> 1. update the interface behavior, so that it is more responsive to users' actions

• when a new item is selected, alert the user

• change the 'item in focus' box so that it is clear it is not editable

• change the java applet into a smoother technology (e.g., flash)

• update the section titles', buttons and pathway descriptions in order to make them easier to understand also for a neophyte

2. add a *generic description* pathway, instead of the 'instance-info' button

3. add a pathway that connects directly instances of 'person' to instances of 'ideas', e.g., something called *list of his/her intellectual products*

8.6 Conclusion

In this section we have presented the design and results of a user-evaluation study of PhiloSurfical, the ontology based tool described in chapter 6. In particular, we gathered feedback from a set of users through the recording of their interactions with the software. The purpose of this experiment was to test the various tabs' functionalities and ease of use, by asking volunteers to accomplish a number of predefined tasks.

We presented the data obtained, and then discussed their significance with respect to the functionalities we originally intended the prototype to support.

In general, the outcome of the experiment was positive for that regards the overall usability and appearance of the tool, but several problems also emerged which require some action on our side. In particular, these can be organized under three broad topics: the user interface, the ontology and the pathways.

In conclusion, the experiment produced a set of requirements which will be used during the development of the next version of the PhiloSurfical tool.

Furthermore, a very interesting follow-up to this initial usability study would be a more pedagogically-oriented task-based evaluation. Despite the numerous difficulties normally associated with the process of measuring 'learning' (Devedzic, 2003), we could still attempt to evaluate how much PhiloSurfical could support learning about philosophy by comparing the performance of a group of tool's users with a control group. In particular, this type of experiment will help us evaluate the learning strategy underlying the system - that is, the idea that learning is facilitated by a narrative-inspired browsing functionality. Such an experiment could be designed as follows:

- a first group of people would have access to the software only, while a second group would have a paper copy of Wittgenstein's work, plus a number of philosophical reference works (e.g., dictionaries, encyclopedias);

- both groups would receive instructions to understand a number of complex relationships among philosophical topics of our choice; for example, by presenting them with a series of tasks similar to number 4 and 5 above (section 8.3.1). Such tasks must involve more than just a look-up type of activity; they should imply the correlation and synthesis of various information that may be found in isolated form, or in different contexts.

- in order to measure their performance, respondents must be given questionnaires to fill out at the end of the experiment; these questionnaires would contain reference to the complex relationships they were supposed

to discover (e.g., 'what philosophical positions resemble a classic atomistic stance in contemporary Germany').

- at the end, a comparison of the two groups' results could give us an idea of which one found it easier to 'navigate' in a philosophical conceptual space, and to construct meaningful representations of this world.

As previously mentioned, we decided that in order to perform a task-based evaluation PhiloSurfical must undergo various refinements. First of all, various improvements concerning the general usability of the website (as emerged in the evaluation study presented in this chapter); second, cleaning up and refining the information present in the knowledge base, so that it would not contains errors (such as names misspelled during the automatic information extraction phase) and it would be more strictly and exhaustively focused around one or two philosophical topics only (e.g., british analytic philosophy). For these reasons, we left the task-evaluation study to future work.