Assessing the User Satisfaction with an Ontology Engineering Tool Based on Social Processes

Ioana Ciuciu and Christophe Debruyne

Semantics Technology and Applications Research Laboratory
Vrije Universiteit Brussel, Belgium
{iciuciu,chrdebru}@vub.ac.be

Abstract. This study discusses one of the three measures defined within the usability testing, namely the user satisfaction, when evaluating an ontology engineering tool based on social processes. The motivation of our focus lays in the fact that being driven by communities through social interactions, the ontology engineering process depends on what the user does, sees and feels when using the system. The evaluation criteria proposed here are therefore developed by looking at the people involved, the processes and their outcomes, mostly taking into account the user experience, in an approach that goes beyond usability. The paper identifies the problems the users encounter when using the system, both at a technical level and psychometric level. A set of recommendations is proposed in order to overcome these problems and to improve the user experience with the system.

Keywords: usability engineering, user satisfaction, hybrid ontology engineering, social processes, community, human-computer interaction, socio-technical systems theory.

1 Introduction

The present study is part of the community evaluation task of the Open Semantic Cloud for Brussels (OSCB[1]) project. OSCB aims at providing a cloud of structured data for the city of Brussels and a platform for publishing and consuming it. Communities work collaboratively on the platform to unlock their data which in turn can be used by other parties to create services for users around the data. The OSCB platform is intended to enable communities of users to represent their knowledge using natural language. The knowledge is further captured in an ontology which is used to annotate data and resources on the web. The ontology is grounded in natural language. The ontology creation is a community effort and is supported by the OSCB knowledge-engineering platform. Note that the multilingual aspect is taken into account in the platform for the particular case of Brussels (i.e. data available in English, French and Dutch).

[1] http://www.oscb.be/

P. Herrero et al. (Eds.): OTM 2012 Workshops, LNCS 7567, pp. 242–251, 2012.

As already mentioned, the aim of this study is to evaluate the ontology-engineering platform of OSCB from the usability engineering point of view. However, our aim is to go beyond usability and to analyze the user experience with the system from a socio-technical perspective: what the user does, sees, and feels when using the platform. The user feedback will be constructively used in order to improve the next iteration of the platform.

This paper reports on the user satisfaction with the OSCB ontology-engineering platform. Based on that, we identify the main (usability) problems and draw valuable conclusions and recommendations for improvement.

The rest of the paper is organized as follows: Section 2 constitutes the paper background. The usability (socio-technical) test design is described in Section 3. Section 4 reports on the results and presents some recommendations for improvement. Section 5 is the related work of the paper. Section 6 concludes and presents the future work of this research.

2 Background

2.1 The Post-study System Usability Questionnaire (PSSUQ)

Satisfaction was measured using the standardized Post-Study System Usability Questionnaire (PSSUQ [1,2]) developed by IBM. PSSUQ originally consisted of 19 questions, each question being a statement about the usability of the system. Participants need to answer each statement using a Likert scale of 7 points, where 1 indicates that the user "strongly agrees" with the statement whilst 7 indicates that the user "strongly disagrees" with it. PSSUQ is based on a comprehensive psychometric analysis, providing scales for three sub-factors, namely: (1) system usefulness; (2) information quality; and (3) interface quality. The short (and most recent) version of PSSUQ, illustrated in Table 1, was used, in order to save study time.

Table 1. PSSUQ – short version [3]

Item	Item Text
Q1	Overall, I am satisfied with how easy it is to use this system.
Q2	It was simple to use this system.
Q3	I was able to complete the tasks and scenarios quickly using this system.
Q4	I felt comfortable using this system.
Q5	It was easy to learn to use this system.
Q6	I believe I could become productive quickly using this system.
Q7	The system gave error messages that clearly told me how to fix problems.
Q8	Whenever I made a mistake using the system, I could recover easily and quickly.
Q9	The information (such as on-line help, on-screen messages and other documentation) provided with this system was clear.

Table 1. (*continued*)

Item	Item Text
Q10	It was easy to find the information I needed.
Q11	The information was effective in helping me complete the tasks and scenarios.
Q12	The organization of information on the system screens was clear.
Q13	The interface of this system was pleasant.
Q14	I liked using the interface of this system.
Q15	This system has all the functions and capabilities I expect it to have.
Q16	Overall, I am satisfied with this system.
SysUse = Average Items 1 through 6 *IntQual* = Average Items 13 through 15 *InfoQual* = Average Items 7 through 12 *Overall* = Average Items 1 though 16	
Table notes: *SysUse = system usefulness; InfoQual = information quality; IntQual = interface quality; Scores can range from 1 (strongly agree) to 7 (strongly disagree), with lower scores better than the higher scores.*	

In this study, we use PSSUQ in order to measure the user satisfaction when dealing with an ontology-engineering tool based on social processes, which is described in the following section. The reason for choosing PSSUQ for this study is mainly the rich information it provides, with little effort from the user, and the extensive IBM documentation and experience for the statistics it can provide. Besides the 16 items in the test, the test participants can make comments and elaborate on their answers. Based on these comments, we will draw conclusions and try to provide recommendations for improving the human-system interaction.

2.2 Grounding Ontologies with Social Processes (GOSPL)

GOSPL [6] is a method and its supporting tool[2] for collaborative hybrid ontology engineering. It supports communities of stakeholder in collaboratively achieving an approximation of the domain to support their semantic interoperability requirements. Hybrid ontologies are ontologies where concepts are both described informally in natural language by means of glosses for high level reasoning between the community members and formally suitable for machine reasoning and data annotation. Concepts are formally described by means of lexons [8], expressing plausible binary relations that hold within a community and expressed in natural language. An example of a lexon is ⟨Cultural Domain, Concert, is a, subsumes, Event⟩, which states that – in the Cultural Domain community- the concept referred to with term "Concerts" plays the role of "is a" on the concept with term "Event" and the concept with term "Event" plays the role of "subsumes" on the concept with term "Concert". A lexon should – ideally – result in two coherent sentences when read in both directions, but assuring this quality is the responsibility of the community. The community that wishes to interoperate formulates semantic interoperability requirements and the social processes

[2] For access to a running instance of GOSPL, please contact the authors.

within that community leading to closer approximations of their domain are driven by those glosses. Meaning agreements across communities reside at two levels, the glossary and the formal descriptions. This allows agreements within the formal and informal part of the hybrid ontology to evolve asynchronically. Ideally, if two communities agree that two descriptions refer to the same concept, the labels associated with those descriptions should be considered equal as well; this is called the glossary consistency principle [6]. All social processes are stored on the platform, adding an additional layer of traceability; not only information is kept about changes in the ontology, the whole discussion leading to this decision is also stored for future reference.

Some of the social processes in the GOSPL methodology include a request to: remove gloss from lexon or term, change gloss of lexon or term, add gloss to lexon or term, change super-type of term, remove gloss-equivalence, add lexon, remove lexon, remove constraint, add synonym, add constraint, change role hierarchy, remove synonym, and add gloss-equivalence. Users are thus able to propose adding, removing, and (sometimes, depending on the subject) change parts of the hybrid ontology. Terms are labels referring to concepts and lexons are binary relations between those terms. Users can create a hierarchy of terms and roles and also describe terms and lexons with a gloss. Synonyms are agreements that two terms refer to the same concept, and gloss-equivalences are agreements that two descriptions refer to the same concept.

Fig. 1. summarizes the different processes in GOSPL [6]. Starting from co-evolving communities and requirements, the informal descriptions of key terms have to be gathered before formally describing those concepts. Communities define the semantic interoperability requirements, out of which a set of key terms is identified. Those terms need to be informally described before the formal description can be added. In order for a lexon to be entered, at least one of the terms needs to be articulated. The terms and roles in lexons can be constrained and the community can then commit to the hybrid ontology by annotating an individual application symbols with a constrained subset of the lexons. At the same time, communities can interact to agree on the equivalence of glosses and the synonymy of terms. Committing to the ontology allows for the data to be explored by other agents via that ontology. Commitments also enable the community to re-interpret the ontology with its extension (i.e. the instances in each annotated system). This will trigger new social processes that lead to a better approximation of the domain, as the community is able to explore the increasingly annotated data, e.g., by formulating queries. We will not go into detail about the commitments, but a description of one such language can be found in [9].

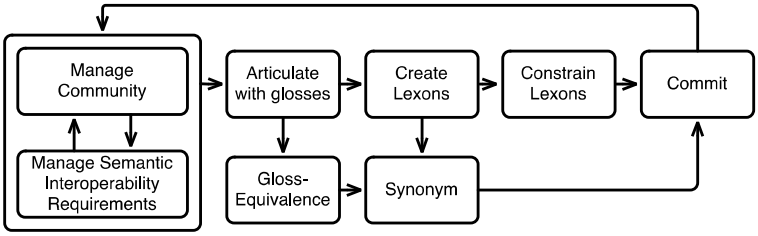

Fig. 1. The GOSPL methodology

Fig. 2 shows a screenshot of the lexons and constraints in the Event Community, which aims to provide a high level, light weight descriptions of events at a particular location and happening at a certain time. This description will be used by more specialized community to "subclass" the concept of Event. Those facts and constraints are the results of discussions between the community members. The screenshot also shows that users have access to the glossary and they can obtain an OWL implementation of the ontology. Indeed, the aim of GOSPL is not to come up with yet another ontology language, but rather use a formalism closer to natural language for ontology engineering out of which other formalisms can be distilled for immediate deployment with other Semantic Web technologies.

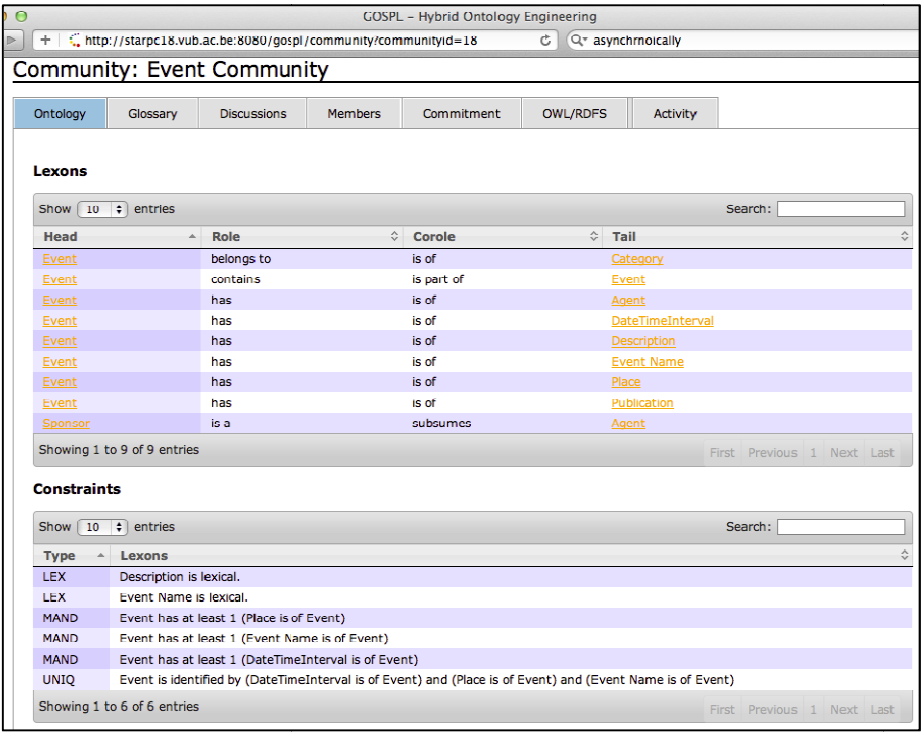

Fig. 2. The GOSPL collaborative ontology-engineering platform. Here, we see lexons and constraints on the formal part of the hybrid ontology. These were the results of social processes within that particular community

3 Test Design

The user satisfaction when interacting with the GOSPL ontology-engineering tool was assessed within a wider usability testing experiment undertaken with a group of master student volunteers of a course on ontology engineering. The goal of the test is to evaluate the usability of GOSPL in two dimensions: *formative* and *summative*,

from the user satisfaction point of view. The formative usability testing aims at identifying the usability problems of the tool. The summative UT consists of a series of measurements (e.g. effectiveness, efficiency, satisfaction) which are performed in order to compare the usability results against a set of predefined objectives.

The objective of the experiment is to create a prototype ontology capturing the (shared) concepts and relations of two applications involving cultural events (e.g. concerts, exhibitions). One information system (IS) is developed by the experiment participants and one application whose database schema and data is provided to them. Both applications are portals. The objectives identified in the test are thus: (1) the ontology creation and (2) the annotation (of the IS and the existing database) with the ontology, together with their subsequent subtasks:

- Propose discussion (the social processes defined in Section 3.1)
- Discuss (vote);
- Conclude (accept/reject discussion);
- Create a community;
- Manage a community (add/delete members and roles);
- Use the ontology to annotate existing information systems.

The satisfaction test was undertaken by a group of fifteen volunteers, divided in four subgroups of three to four people. The volunteers were recruited from university students with the distribution of gender, age, education and occupation very close in the four groups. The participants are also grouped by the community in which they get involved based on their application: the 'Date and Time' community, the 'Ticket' community, the 'Venue' community, etc.

The overall usability testing was carried out both implicitly by analyzing the data logs and the user-system interactions and explicitly, by collecting the user feedback via several questionnaires. The outcome of the experiment highlights three aspects of the evaluation: 1) effectiveness; 2) efficiency and 3) satisfaction [3]. Following the recommendations in [10] we have developed the evaluation criteria looking at the people involved, the processes and their outcomes.

The purpose of this study is to assess the user satisfaction with the system. The results are reported in the following section.

4 User Satisfaction. Results and Recommendations

4.1 Summative User Satisfaction

The results delivered by the PSSUQ questionnaire are as follows: the overall (average) user satisfaction shows a value of 3.4, the system usefulness 3.1, information quality 3.9 and interface quality 3.4 (see Table 2). There were four groups of volunteers involved in the study, denoted as: A, B, C and D. The group the most satisfied overall is group C, the group the most satisfied with the system is the same group C, the group the most satisfied with the information quality is group B and the group the most satisfied with the interface quality is group A. The overall satisfaction ranking

over the four groups is therefore: group C (most satisfied), group B, group D and group A (least satisfied).

Table 2. Summative user satisfaction

Metric	A	B	C	D	Average
SysUse	3.4	3	**2.8**	3.1	3.1
InfoQual	4.1	**3.8**	3.9	3.9	3.9
IntQual	**3.1**	3.6	3.3	3.6	3.4
Overall	3.6	3.4	**3.3**	3.5	3.4

We have compared these results with the system logs indicating the number of user-system interactions per group, in order to justify the differences on the satisfaction levels per group. The logs show that the average number of interactions per group corresponds to the overall satisfaction per group: 1) group A - 94 interactions; 2) group B - 270 interactions; 3) group C - 350 interactions; and 4) group D - 176 interactions. These results are also correlated with the quality of work of each group, reflected also by the final project grade. It is curious, however, to observe that the group which has used the interface the least often (group A) is the most satisfied with the interface quality. Overall, they are the least satisfied with the system.

4.2 Formative User Satisfaction

The problems identified by the users in the comments section of each item are illustrated in Table 3.

Table 3. Formative user satisfaction

	Usability Problem	Nature	No of reports
1.	The (error) messages displayed by the system were often not clear to the user. There was in general no online help or documentation available	InfoQual	6
2.	There is no “undo” or “edit” option available	InfoQual IntQual	5
3.	No (top menu) link to the current community in the discussion page	IntQual InfoQual	5
4.	It took a while to understand how the system works	SysUse	3
5.	Sometimes, listing items in the dynamic tables didn’t go well when after returning to a page it displayed the first item again	IntQual	2
6.	There was no “delete” option for the communities who “died” during the process	InfoQual IntQual	1
7.	The user name is not clear (just email addresses appear)	IntQual	1
8.	Sometimes, more clicking necessary that one would expect (e.g. when browsing through several discussions)	SysUse	1

The problem they faced most often was linked to the absence of intelligible (error) messages and on-line help/documentation. Other often cited problems are related to the lack of "edit" and "undo" options; also, the users mention the need for a better organization and links regarding the (current and visited) communities. Other problems include clarification of the user names, the "back" button or the missing request to add gloss in the list box.

4.3 Recommendations for Improvement

Taking the satisfaction results obtained from PSSUQ and the user comments, we drive the following conclusions: out of the three sub-factors identified by PSSUQ, the *system usefulness* measure performed best (3.1); the *information quality* of the system is the sub-factor that needs most improvement (3.9). This also corresponds to the (InfoQuality-related) problems the most cited by the users: usability problems 1, 2 and 3 in Table 3. Therefore, summative and formative usability testing deliver similar results when it comes to user satisfaction in this study. An improvement regarding problem 1 is to introduce more *information support* for the user in the form of intelligible error messages or documentation. The students recommend an option to activate/disable these messages.

Regarding the *"delete"* and *"edit"* options, the students were explained that we were reluctant for allowing editing of posts/comments (e.g. to prevent abuse). Most of them would be content with a feature that allows a post to be edited within a number of seconds.

Regarding the problems related to *communities*, the users mention that there is no link to the current community in the discussion page. Such a link exists, but the users didn't notice it. Even though it was styled as a link with the website's style sheet, the users seemed to have overlooked it. There was also a wish from the users to delete communities, in particular the communities that became obsolete (or – as the users put it – "dead") as the different communities evolved. Even though they understood that even those communities might once again become active, they would be happy to be able to "filter" the dead communities from the list and toggle that filter.

Other identified improvements are as follows: the *organization of the information* could be improved (e.g. show more entries by default – the default number of items shown in a table is 10 and users wish to augment this number for caret browsing, more clearly show in which community you are at a certain moment, the visited communities, etc.); *notifications* from the system could be useful and/or a list of changes after the last visit; showing the *user name*, not just email addresses; more *visibility* with respect to the *visited communities* would be very handy (more forum-like).

5 Related Work

Usability is defined by the ISO-9241 standard [11] as the *effectiveness*, *efficiency* and *satisfaction* with which specified users can achieve specified goals in particular

environments. Usability is a key factor in making the (computer) systems easy to learn and to use. Usability testing has been extensively studied and applied by Lewis [3] at IBM Software Group. The results of the usability testing improve the design of a system by evaluating the organization, presentation and interactivity of the system interface.

The subject of this study is the *user satisfaction*, evaluated while using an ontology-engineering system. A standard instrument to assess the user satisfaction is the *Post-Study System Usability Questionnaire* (PSSUQ). PSSUQ was developed for scenario-based usability evaluation at IBM [1]. The environment used was enterprise-wide and networked office application suites. A follow up study by IBM [2] was performed in the domain of speech recognition [5] using data from five years of usability studies. The follow up produced similar psychometric properties as the original survey. Fruhling [4] validates these results of the PSSUQ instrument and assesses its adaptability to other domains, such as telemedicine.

Socio-Technical Systems Theory (STST) has been widely applied in the domain of information systems implementation [12]. Even though older than 50 years, STST witnesses a revival in the information age. Eason [13] stresses that the important aspect about STST is that it provides the means of understanding the way in which people at work collaborate and use technical artifacts to achieve their tasks. It has been argued that STST determined the movement towards a user-centered approach [14].

Nowadays, we witness the tendency to combine the socio-technical concepts and usability engineering in the design of information systems [15]. In [10], the author questions: "*Is usability enough?*", arguing that the existing usability measures are not capturing all that is of interest in order to design more innovative systems. According to that study, the user experience exists at three levels: the process, the outcome and the affect. More precisely, one needs to analyze what the user does, what the user attains and what the user feels. Not all of these aspects have a direct and clear way of being measured. Taking into account the emotional and affective factors would help to understand the human-system dynamics more efficiently.

6 Conclusion and Future Work

We conclude the paper by reminding that the aim of the study was to assess the user satisfaction with the GOSPL ontology-engineering tool, which is part of a larger, multi-lingual knowledge-engineering platform. The standard PSSUQ IBM questionnaire was used as instrument for the evaluation of the tool with 15 master students in Open Information Systems. The results show that the system performs best regarding the system usefulness sub-factor (3.1 on a scale from 1 to 7 were 1 is the best result and 7 is the worst result) and that it needs some improvements regarding the way the information and interface are structured and presented to the user for interaction.

The study not only takes into account the classic usability testing measures, but also brings together the usability measures and the socio-technical systems theory principles in a user-centric approach.

A reiteration of the test with an improved version of the interface, after taking into account the recommendations from this paper is work in progress. Future work includes testing the user satisfaction (and the usability in general) from a socio-technical systems theory point of view with users from various domains, different than students, preferably within one the OSCB project use cases.

Acknowledgments. This study is supported by the INNOViris Open Semantic Cloud for Brussels (OSCB) project, financed by the Brussels Capital Region.

References

1. Lewis, J.R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Technical Report 54.786 (1993)
2. Lewis, J.R.: Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. J. of Human-Computer Interaction 14(3), 463–488 (2002)
3. Lewis, J.R.: Usability testing. In: Handbook of Human Factors and Ergonomics, 4th edn., pp. 1267–1312. John Wiley, New York (2012)
4. Fruhling, A., Lee, S.: Assessing the Reliability, Validity and Adaptability of PSSUQ. In: Americas Conference on Information Systems, Omaha, USA (2005)
5. Lewis, J.R.: Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. J. of Human-Computer Interaction 14(3), 463–488 (2002)
6. Debruyne, C., Meersman, R.: GOSPL: A Method and Tool for Fact-Oriented Hybrid Ontology Engineering. In: Morzy, T., Haerder, T., Wrembel, R. (eds.) ADBIS 2012. LNCS, vol. 7503, pp. 153–166. Springer, Heidelberg (2012)
7. Debruyne, C., Meersman, R.: Semantic Interoperation of Information Systems by Evolving Ontologies through Formalized Social Processes. In: Eder, J., Bielikova, M., Tjoa, A.M. (eds.) ADBIS 2011. LNCS, vol. 6909, pp. 444–459. Springer, Heidelberg (2011)
8. Meersman, R.: The use of lexicons and other computer-linguistic tools in semantics, design and cooperation of database systems. In: CODAS 1999, pp. 1–14. Springer (1999)
9. Verheyden, P., De Bo, J., Meersman, R.: Semantically Unlocking Database Content Through Ontology-Based Mediation. In: Bussler, C.J., Tannen, V., Fundulaki, I. (eds.) SWDB 2004. LNCS, vol. 3372, pp. 109–126. Springer, Heidelberg (2005)
10. Dillon, A.: Beyond usability: process, outcome and affect in human-computer interactions. Canadian Journal of Library and Information Science 26(4), 57–69 (2001)
11. ISO. ISO 9241-11:1998: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. Source (1998), `http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883`
12. Eason, K.: Information Technology and Organizational Change. Taylor and Francis, London (1988)
13. Eason, K.: Sociotechnical systems theory in the 21st Century: another half-filled glass?, Sense in Social Science: A collection of essays in honour of Dr. Lisl Klein, Desmond Graves, Broughton, pp. 123–134 (2008)
14. Baecker, R., Grudin, J., Buxton, W., Greenberg, S.: Readings in Human-Computer Interaction. In: Toward the Year 2000. Morgan-Kaufman, New York (1995)
15. Dillon, A.: Group dynamics meet cognition: combining socio-technical concepts and usability engineering in the design of information systems. In: The New Socio Tech: Graffitti on the Long Wall, pp. 119–126. Springer, London (2000)